VOL. 10 No. 1, pp. 1-36 JULY 1948

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
THE PHILIPS INDUSTRIES

EDITED BY THE RESEARCH LABORATORY OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN, EINDHOVEN, NETHERLANDS.

# X-RAY SPECTROMETER WITH GEIGER COUNTER FOR MEASURING POWDER DIFFRACTION PATTERNS

by J. BLEEKSMA *), G. KLOOS *) and H. J. DI GIOVANNI *).

539.26:535.322.1:
539.16.08

With the spectrometer described in this paper, which was engineered by North American Philips Co., Inc., along the lines indicated by H. Friedman, X-ray diffraction patterns are traversed by a Geiger counter tube detector instead of being recorded in the conventional fashion on photographic film. The Geiger counter tube, also developed by H. Friedman in a form especially adapted to the measurement of X-ray intensities, is applied in a well known focusing arrangement, making use of a large flat specimen (dimensions of active area, e.g., 1 × 2 cm). Due to the extremely high sensitivity of the counter tube, to the high apparent X-ray "brightness" of the focal spot which is used at a glancing angle of only 1.5°, and to the large specimen surface contributing to the diffraction intensity, an X-ray tube rating as low as 125 W is sufficient. The measurement of the intensity of the diffraction lines is effected, *a*) by counting the current pulses produced in the Geiger tube by the entering X-ray quanta, a scaling circuit making this possible at radiation intensities of several thousand quanta per second at the tube window, or *b*) by measuring, after amplification, the mean value of the current through the Geiger tube by means of a milliammeter; the milliammeter reading can be automatically recorded while the spectrum is being scanned, thus producing a chart which immediately indicates the relative line intensity as a function of the diffraction angle. The scanning speed, as well as the averaging time for the mean current measurement, can be chosen between wide limits. The accuracy of the determination of diffraction angles and relative intensities, and the resolving power for adjacent diffraction lines are better than with the commonly used photographic methods, under comparable electric and geometric conditions. Moreover the Geiger counter spectrometer offers additional distinct advantages for applications in which the accurate measurement of only a few lines is desired. A more detailed account of the special techniques and possibilities of the instrument will be given later.

The use of X-ray diffraction technique for chemical analysis, for metallurgical investigations, etc. has become rather commonplace in the past two decades. From an interesting laboratory phenomenon, X-ray diffraction has thus become an invaluable tool for industry. Many examples of its industrial application have been given in papers in this Review [1]).

Actually, it has been apparent for a number of years that the possible fields of application of the method are much wider than those covered at the moment. The limitations on the use of X-ray diffraction are generally not set by the results of which the method is capable, but more commonly by the time and cost involved in its practical adaptation to a particular type of problem.

In the past few years a very promising line has been opened in the field of X-ray diffraction techniques by the development of a new instrument, the Geiger counter spectrometer. By means of this instrument the desired information can in certain cases be obtained in much less time than with the conventional photographic procedure. Moreover, routine measurements can be made by relatively unskilled operators, while at the same time the accuracy of the results is in general better than with commonly used photographic methods. Although it is not to be supposed — because of certain limitations of the new method — that the photographic techniques will be superseded, we may

*) North American Philips Co., Inc., New York, N.Y.

1) Cf. the series of articles by W. G. Burgers in Philips Techn. Rev. **1** (1936) and **2** (1937), the summarizing article in **5**, 161, 1940, and another series by J. F. H. Custers in Philips Techn. Rev. **7** (1942).

safely expect that a considerable increase in the industrial use of X-ray diffraction will be brought about by this instrument.

In this paper a succinct description of the principle and design of the Geiger counter spectrometer is given. Limits of accuracy, techniques of practical use and a number of special applications will be described in a second paper.

**Brief outline of X-ray diffraction; photographic method**

The diffraction of an X-ray beam striking a crystal can be described as a reflection, one set of the lattice planes in which the atoms are arranged acting as a mirror (*fig. 1*), under the condition that the angle $\Theta$ of "reflection", the X-ray wavelength $\lambda$ and the spacing $d$ between successive parallel lattice planes of the set conform to the well-known Bragg equation:

$$n\lambda = 2\,d \sin \Theta,$$

where $n$ is an integer indicating the order of the diffraction.

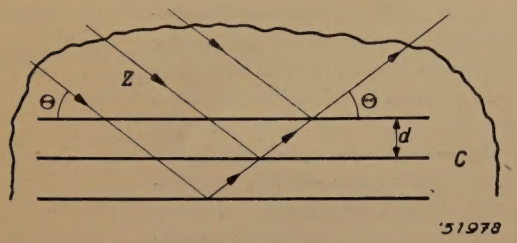



Fig. 1. Diffraction of a beam of X-rays $Z$ impingent on a crystal $C$ can be described as a reflection at a set of lattice planes, conforming to the Bragg relation between the angle $\Theta$, the wavelength $\lambda$ of the rays and the spacing $d$ of the parallel atomic planes.

In a single crystal of a material a series of different sets of lattice planes can be distinguished, every set possessing a specific "planar spacing value" $d$ and a specific orientation within the crystal. Using X-rays of one single wavelength, reflection at a given lattice plane will occur only if the crystal has an appropriate position with respect to the beam, the required angle between the beam and the plane being fixed ($\Theta$). The intensity of the reflected beam depends on the arrangement, kind and density of the atoms in the plane, and on the perfection of the crystal, as well as on the absorption of the X-rays in the material.

In most practical applications the substance under examination is not a single crystal, but a polycrystalline material, e.g. a powder. Assuming the tiny crystals of the powder to be oriented entirely at random, there will always be a number of crystals showing the orientation required for a specific lattice plane to act as a mirror and hence giving rise to a diffraction of X-rays in the corresponding direction $\Theta$. The angle $\Theta$ being fixed only with respect to the incident beam, and rotational symmetry prevailing because of the random orientation of the crystals, the diffracted X-rays form a cone with included angle $4\Theta$ around the primary beam. The entire diffracted radiation from the powder specimen consists of a series of such coaxial cones, belonging to different sets of lattice planes in the elementary crystal (*fig. 2*).

A hypothetical sphere arranged around the powder specimen is intersected by the cones in a series of concentric rings, the central spot being formed by the primary beam. By laying a narrow strip of photographic film in a circle around the specimen, a small part of every ring is recorded on each side of

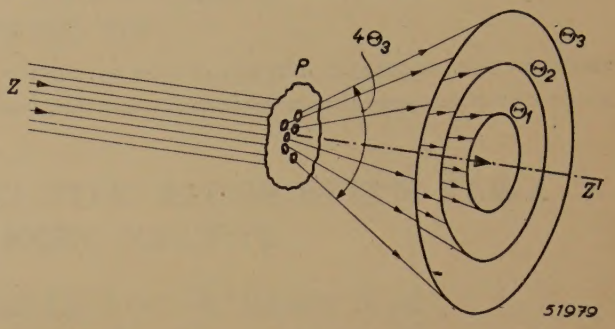



Fig. 2. When a polycrystalline specimen $P$ containing crystallites oriented at random is placed in a monochromatic X-ray beam $Z$, the rays diffracted at a Bragg angle $\Theta$ by the suitably oriented crystallites form a cone with included angle $4\Theta$ around the direction $Z'$ of the primary beam. Different lattice planes (spacing values $d$) give rise to different cones.

the central spot. In this way the well-known diffraction pattern (Debye-Scherrer photograph) of the specimen substance is obtained (*fig. 3*).

After this brief recapitulation of the principles of X-ray diffraction, the photographic method for the measurement of diffraction patterns requires little further explanation. The arrangement generally used is shown in *fig. 4*. The specimen is formed, for example, by a small cylindrical tube filled with the powder under examination and placed along the axis of a cylindrical camera. The X-rays emanating from the anode of an X-ray tube are collimated by means of a pair of pin-hole apertures of say 0.5 mm width; the fine pencil of X-rays thus obtained enters the camera and impinges on the specimen. A film is laid against the whole inner circumference of the cylindrical wall of the camera. With ordinary equipment the exposure times necessary for obtaining a diffraction pattern of sufficient density vary between 15 minutes and more than 10 hours. In cases where the exact knowledge of relative

Fig. 3. Diffraction pattern of a specimen obtained on a photographic film. The "lines" (rings) are the intersections of the film with the series of cones shown in fig. 2. (The shadow along the lower left hand edge of the pattern is produced by a thin nickel filter, offering a means of identifying undesired K$\beta$-lines of the copper radiation; cf. footnote [8]).)

intensities of the diffraction lines is required (e.g., for quantitative analysis of mixtures), the density of the film pattern must be measured by means of a microphotometer.

This is not the place to dwell on the various types of information which can be read from the position, the intensity and the shape of the lines in the diffraction pattern [2]). We confine

[2]) Cf. e.g. W. G. Burgers, Philips Techn. Rev. 5, 157, 1940.

ourselves to mentioning the identification of substances by means of the so-called "fingerprint" method. The crystal lattice being different for all substances, each having its specific planar spacing values and corresponding diffraction intensities, no two patterns are exactly alike. Hence, a sub-

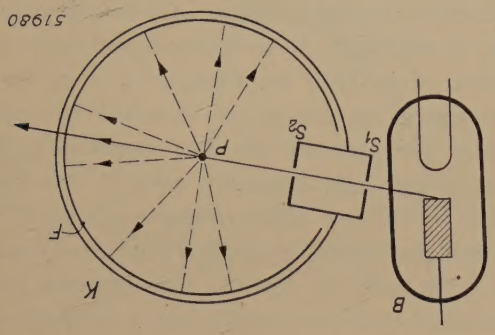



Fig. 4. For the photographic recording of powder diffraction patterns, a cylindrical camera is commonly used. The X-rays emitted by the tube $B$ are collimated in a narrow beam by pin-hole apertures $S_1$, $S_2$, impinge on the thin cylindrical specimen $P$ at the centre of the camera $K$, and are diffracted onto the film strip $F$.

stance can be unambiguously identified by the spacing values derived from a number of lines in its diffraction pattern. The spacing values of most of the commonly occurring substances have been listed in a catalog, compiled by the American Society for Testing Materials and indexed according to the three strongest diffraction lines (the "fingerprints" of each substance). Such a catalog offers a simple means of chemical identification and analysis of mixtures.

## Principles of the Geiger counter spectrometer

In the photographic method for investigating diffraction patterns, local radiation intensities are measured by the blackening of a photographic film. There are other, and in some cases, more efficient means for radiation measurements; one of these is the Geiger counter tube, *fig. 5*. When a single X-ray quantum is absorbed in such a tube, it sets free a photo-electron which gives rise to a short discharge through the gas filling between the wire anode and the surrounding cylindrical cathode. By counting the short current pulses through the Geiger tube produced by the absorbed quanta, the rate of the arrival of quanta, i.e. the radiation intensity, is measured. We shall give below a more detailed account of the properties of the Geiger tube and of its adaptation to our special purpose.

While the photographic film registers all diffracted beams simultaneously, the Geiger tube can measure the mean intensity at only one spot at a time, viz. at the instantaneous position of the window through which the X-rays enter the tube. Therefore, the Geiger tube must be moved along the circle on which the film would have been mounted in the photographic method, and the intensities at different diffraction angles must be measured one after the other. The simultaneous recording of the whole spectrum of diffraction lines would seem to be an inherent advantage of the film — and in fact for some purposes it is — but in general this advantage is outweighed by the greater sensitivity of the Geiger tube: to obtain a sufficient blackening of the film, with a wavelength of 1.5 Å, an incident radiation energy of about 2 erg per cm$^2$ is necessary, while one single line can be recorded by the Geiger counter using only say $4 \times 10^{-4}$ erg/cm$^2$ (for a precision of about 5% in the intensity measurement).

A local radiation detector, such as the Geiger tube, is used to best advantage in conjunction with a focusing arrangement; cf. *fig. 6*. Instead of the small cylindrical specimen previously mentioned, a large flat specimen with dimensions of perhaps $1 \times 2$ cm is used. A divergent X-ray beam, emerging from the slit $A$, irradiates this specimen at an angle $\Theta$. Considering only the crystallites (of appropriate orientation) lying at the surface of the specimen, it is seen that the rays reflected at the angle $\Theta$ from all points of the specimen very nearly converge to a single line at point $B$ in fig. 6; at this point the entrance window of the Geiger tube is placed. The "focusing" at $B$ would be exact if the specimen were curved to conform to the circle


Fig. 5. A Geiger counter tube for X-ray measurements. Along the axis of a cylindrical electrode which forms part of the tube wall, a straight wire electrode is placed. Between these electrodes a high d.c. voltage is applied, the wire acting as the anode.

drawn through points $A$, $B$ and $P$; all angles inscribed in this "focusing" circle on the chord joining $A$ and $B$ are equal, and so are the values of angle $\Theta$ for all the paths of incident and reflected beams going through $A$ and $B$. The deviation of the flat specimen from the focusing circle is, within certain limits, not serious, but obviously the specimen must be made to conform as nearly as possible

to the circle, i.e. it must be tangent to this circle in $P$. Therefore, on moving the Geiger tube (i.e. point $B$) on the scanning circle $C$ around the axis P, the specimen must rotate around the same axis at half the angular speed of the Geiger tube, thus always maintaining the mirror position between points $A$ and $B$ which is necessary for the osculating condition mentioned above.

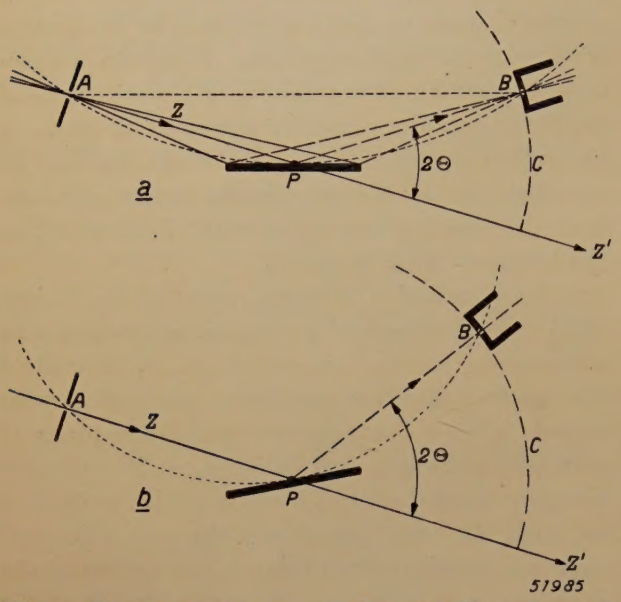



Fig. 6. Focusing arrangment. A divergent X-ray beam $Z$ coming from slit $A$ irradiates the surface of the flat specimen indicated by the heavy line through $P$. All the rays (dotted lines) diffracted at Bragg angle $\Theta$ by suitably oriented crystallites at the surface very nearly converge to a single line, projected at point $B$. Here the receiving slit of the Geiger counter tube is placed. When the counter tube is moved along the circle $C$ around axis $P$, in order to scan subsequent angles $2\Theta$, the flat specimen must rotate around $P$ in such a way that it is always tangent to the focussing circle through $A$, $P$, $B$ (compare $a$ and $b$).

With the conventional non-focusing arrangement with cylindrical specimen, the diffraction lines obtained have a width comparable to the diameter of the specimen. Therefore the specimen (and hence the irradiating pencil of X-rays) must be very narrow in order to obtain sufficient resolution of the lines on the film, although larger specimens and correspondingly wider beams would be desirable for rapid recording of lines. The large flat specimen used in the focusing arrangement receives and reflects about 10 to 20 times as much radiation as the cylindrical specimens used in ordinary photographic technique.

Incidentally, it will be noted that in the focusing arrangement it is presupposed that the diffracted X-rays contributing to the measured intensity originate from suitably oriented crystallites in a very thin specimen layer. A thin specimen can be prepared from most materials by one of several methods. In cases where a fairly thick specimen must be examined, the diffracted energy contributed from crystallites lying at some depth beneath the surface will give rise to a greater line width. This point will be dealt with more extensively in the second article.

The principle outlined in this section, viz. the use of a local radiation detector in a focusing arrangement, is not new. It was developed and applied as early as in 1913 by Bragg [3]), who, however, made use of an ionization chamber as radiation detector. Because of the low sensitivity of this detector and the constant care it requires, the method could not then compete on equal terms with the photographic method and never was used on a large scale. In fact the new impetus given to the method to-day is largely due to the replacement of the ionization chamber by the Geiger counter tube which in the meantime has been developed from a somewhat unfamiliar laboratory instrument, used mainly for cosmic ray experiments, into a rugged and reliable device of utmost sensitivity. The change-over to the Geiger counter was initiated by H. Friedman [4]), who also changed the design of the Geiger tube so as to render it particularly sensitive to the X-rays used for the diffraction measurements. Friedman's ideas were executed in a practical form by the North American Philips Company, Inc., which engineered and produced the spectrometer described in the following section.

## General description of the apparatus

### *The set-up*

The set-up of the apparatus is shown in *figs 7* and *8*. The source of the X-ray beam (focal spot of the X-ray tube) and the entrance slit of the Geiger tube are each spaced 13 cm from the specimen axis. The width of slit $S_1$ controls the divergence of the primary X-ray beam, and may be chosen so that at a Bragg angle of, e.g., $\Theta = 15°$ a $1 \times 2$ cm surface of the specimen is irradiated. The Geiger counter

[3]) W. H. Bragg, Proc. Roy. Soc. A **88**, 428-438, 1913. In this paper the arrangement was applied to single crystals; the application to powder specimens was described by Bragg in Proc. Phys. Soc. **33**, 222-224, 1921.

[4]) H. Friedman, Geiger counter spectrometer for industrial research, Electronics, April 1945, p.p. 132-137. This apparatus, showing most of the features incorporated in the present design, was constructed in the Naval Research Laboratory at Washington, D.C. The present design is in part an outgrowth of machines which Friedman also aided in designing, and which were built at an earlier date by North American Philips Co. and widely used during the war for accurately controlling the orientation of quartz oscillator plates; see W. Parrish and S. G. Gordon, Precise angular control of quartz-cutting by X-rays, Amer. Mineralogist **30**, 326-346, 1945. — An independent development of similar ideas appears to have taken place in Germany in the beginning of the war: R. Lindemann and A. Trost, Das Interferenz-Zählrohr als Hilfsmittel der Feinstrukturforschung mit Röntgenstrahlen, Z. Physik **115**, 456-468, 1940.

tube entrance slit $S_2$ must be narrow in order to obtain a high resolving power for adjacent diffraction lines. On the other hand, more diffracted radiation enters the counter tube, and hence measurements may be speeded up, with a wider slit. As an adequate compromise to cover the ordinary range of requirements, the width of the entrance slit can be adjusted to 0.25, 0.5 or 1.0 mm.

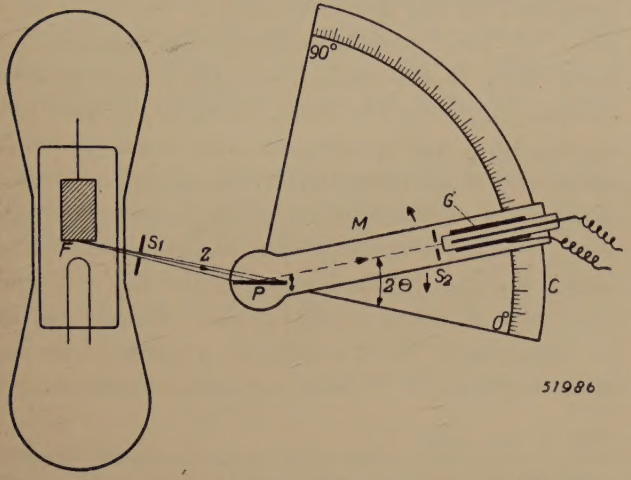



Fig. 7. Set-up of the Geiger counter spectrometer. The divergence of the X-ray beam $Z$ emitted from the small focal spot $F$ is controlled by the slit $S_1$. Slit $S_2$ confines the diffracted beam entering the Geiger counter tube $G$. This tube is mounted on the goniometer arm $M$ and slides along the goniometer circle $C$ which carries a Bragg angle scale from $2\Theta = 0°$ to 90°. The specimen holder $P$ is geared with the goniometer arm $M$ so as to rotate at half the angular speed of the counter tube.

The Geiger counter tube is mounted on a scanning arm pivoted at the specimen axis. The position of the arm corresponds to the angle $2\,\Theta$, i.e. twice the Bragg angle, and is read on a goniometer scale and vernier graduated to 0.01°. The specimen holder is geared with the scanning arm so as to rotate at half the angular speed of the counter in accordance with the focusing condition described above. The mechanism is free of backlash to such a degree that resetting is possible with an accuracy of $\pm$ 0.01°. The exact positioning of the specimen holder and the goniometer scale is accomplished with the aid of a flat single quartz crystal whose surface is cut parallel to a known lattice plane (with precisely known spacing value $d$).

### *The focal spot*

It has been pointed out above that the focal spot of the X-ray tube acts as the geometric origin of the divergent beam falling on the specimen. A previ-

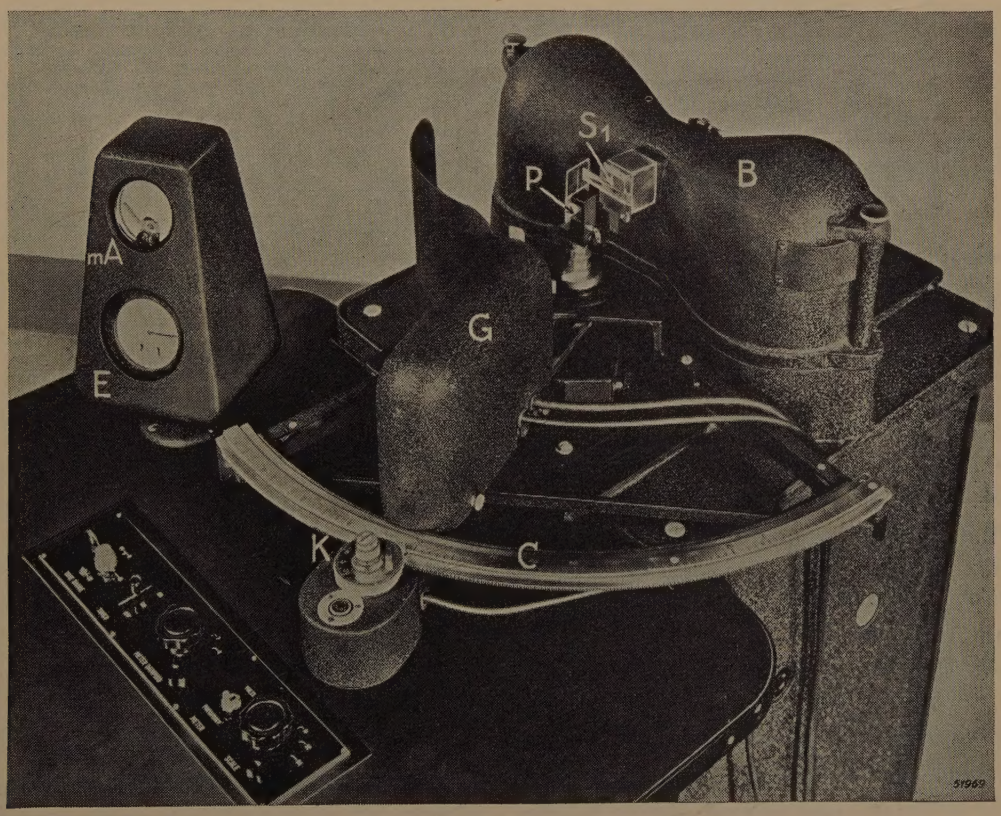



Fig. 8. Photograph of the desk of the Geiger counter spectrometer, showing the arrangement of X-ray tube housing $B$, slit $S_1$, specimen holder $P$, Geiger counter tube housing $G$ and goniometer circle $C$. $K$ is the vernier knob for moving the counter tube over the goniometer scale; $mA$ and $E$ are the instruments for reading the diffraction intensities, described below.

ously used solution consisted in using as the beam origin a narrow "source slit" irradiated by a large focal spot. Obviously the present solution is more economical with regard to the use of the X-ray energy. In this case very small dimensions of the focal spot are required. The condition of sufficient resolving power for most practical purposes allows a source width of 0.25 mm, while the height of the source, being parallel to the "lines" in the diffraction pattern, can be made comparatively large, e.g. 2.5 mm. These required dimensions apply, however, to the apparent focal spot, i.e. the projection of the focal spot in the direction of the useful beam. The real width of the focal spot can be made much larger by having the useful beam oriented obliquely to the anode surface, resulting in an increase in beam intensity [5]). In our case the intensity is 35 times the value in the direction normal to the anode, as the beam includes the exceptionally small angle of 1.5° with the anode; the real width of the focal spot in this case is 7 mm, the apparent width 0.2 mm (*fig. 9*).

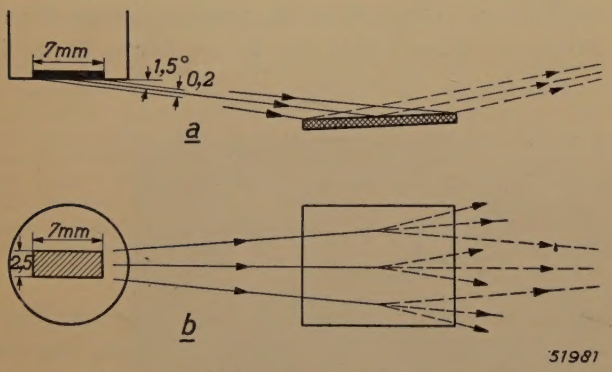



Fig. 9. *a*) The real width of the focal spot is 7 mm, while the "apparent width", due to the very small projection angle of 1.5°, is only 0.2 mm. *b*) The apparent height of the focal spot, being parallel to the diffraction lines (and to the entrance slit of the Geiger counter tube), is equal to the real height, i.e., about 2.5 mm.

As the large height of 2.5 mm for the beam source (focal spot dimensions perpendicular to plane of fig. 7) is only permissible when it is parallel to the diffraction lines, it will be evident that the axis of an X-ray tube of conventional design must lie in the plane traversed by the scanning Geiger tube. Since, for the sake of convenience, the goniometer scale is arranged in a horizontal plane, the X-ray tube must also be mounted horizontally (figs 7 and 8). This entails the limitation of the goniometer scale to an angle $2\Theta = 90°$, as the further movement of the counter tube needed for traversing larger angles is prevented by the X-ray tube housing.

The very small angle between the anode surface and the useful X-ray beam requires extremely careful polishing of the surface, as minute irregularities would otherwise be likely to intercept part of the beam. During the life of X-ray tubes, some roughness of the anode surface is ordinarily produced by the evaporation of material from the hot focal spot. In our case, however, this roughness remains imperceptibly small, owing to the very low specific focal spot loading. In fact, due to the high sensitivity of the detecting device and the large gain in intensity achieved by use of the focusing arrangement, an X-ray tube power as low as 125 W is sufficient. The specific loading of the 2.5 × 7 mm focal spot therefore is only about 7 W/mm², while in high-power water-cooled X-ray tubes specific loads of 60 to 80 W/mm² are quite common.

*The X-ray tube*

The X-ray tube is fed with a.c. and driven at 35 kV (peak value) and 5 to 6 mA (mean value). The small power of 125 W can easily be dissipated by air cooling instead of requiring the usual water cooling — a very welcome simplification of working conditions. The filament current and tube voltage transformers are supplied by a stabiliser. Moreover, means are provided which correct for drift of the filament current during the warm-up period. These precautions are necessary because the diffraction lines of a pattern are not measured simultaneously but successively; as a consequence, the intensity of the primary X-ray beam must be kept constant during the whole scanning time in order to avoid errors in the relative line intensities.

As is usual in X-ray diffraction practice, the X-ray tube is made readily interchangeable to facilitate the selection of a suitable target material for the focal spot. Tubes with molybdenum, copper or iron target are available at present, yielding the characteristic radiation of these elements at 0.7107, 1.5418 and 1.9273 Å respectively (weighted mean value of $K\alpha_1$ and $K\alpha_2$ lines). Thus for practically all specimens a radiation with sufficiently long wavelength can be used to avoid the excitation in the specimen atoms of their own characteristic radiation (fluorescence), which would increase the undesired background intensity of the diffraction pattern.

*Circuit for the intensity measurements*

The current pulses produced in the Geiger counter tube by the arriving X-ray quanta are preamplified in two amplifier stages, then in a third stage the pulses are equalized, i.e., they are given a definite shape independent of the duration and intensity of the discharge in the counter tube, which may vary

[5]) See e.g. Philips Techn. Rev. 3, 261, 1938.

to a certain degree. The rate of the production of pulses, which is used as a measure of the X-ray intensity, can be determined in two ways:

*a*) A mechanical register is provided, similar to those used for registering telephone calls, that is actuated by a relay into which the pulses are fed (*fig. 10*). Because of the inertia of the mechanical register, its resolving power for successive pulses

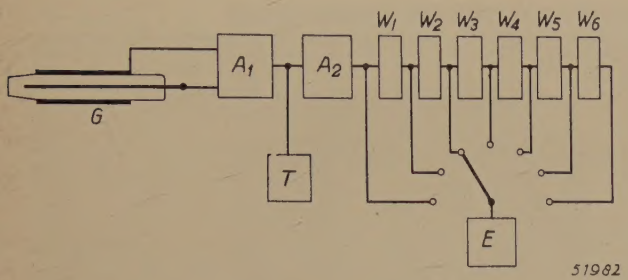



Fig. 10. Schematic circuit for counting the pulses produced in the Geiger counter tube *G*. The pulses are preamplified in $A_1$, then once more amplified and equalized in $A_2$, and counted by the electromechanical register *E* (cf. also fig. 8). At pulse rates greater than about 60 per second, the pulses are scaled down by a number of scale-of-two circuits $W_1 \ldots W_6$, allowing a maximum scaling down factor of 64. The counting interval is controlled by the timer *T*.

is limited to about 60 counts per second. In order to profit by the much higher resolving power of the Geiger tube itself (several thousand pulses per second), a scaling circuit is inserted between the tube and the counting system. This well-known device consists of a number of scale-of-two (flip-flop) circuits connected in series. Every one of these scale-of-two circuits [6]) essentially contains two twin-triodes connected in such a way that only one of the two can be in a conducting state at a time, while an arriving pulse of a given polarity causes both tubes to change their state; therefore, in the output of the circuit only one pulse (of a given polarity) occurs for every two pulses arriving at the input. In our case, six such units are provided. The output of the last scaler unit actuates a relay-tube which passes the impulses on to the register.

For the measurement of high intensities the full scaling down factor of $2^6 = 64$ is used; for smaller intensities intermediate factors can be chosen by omitting a number of the scale-of-two units. Often it will be sufficient to count the pulses received during a period of half a minute or so. The time period over which the counter will operate is controlled by a timer, cutting off the pulse output between timing intervals and allowing for a continuously variable setting up to 64 sec. The settings of 64, 32 and 16 seconds are marked, as with these settings the number of counts registered is directly equal to the number of counts per second when using 6, 5, or 4 scaling-down units respectively.

*b*) The rate of arrival of quanta can also be determined by measuring the mean value of the current flowing at the amplifier output. This current is used to charge a condenser *C* shunted by a resistor *R* (*fig. 11*). The voltage rise on *C* within certain limits is proportional to the number of pulses per second, all pulses yielding identical contributions owing to the equalization stage mentioned above and to proper design of the circuit (such that the voltage rise for each pulse is approximately independent of the voltage already attained on *C*). The leakage current through *R* being proportional to the voltage on *C*, a milliammeter connected in series with *R* indicates directly the X-ray intensity entering the Geiger tube.

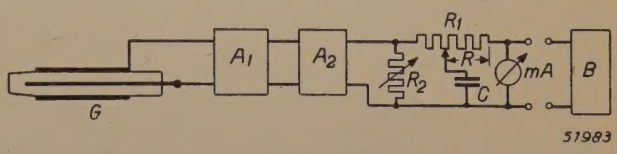



Fig. 11. Circuit for measuring the mean value of the current flowing through the Geiger tube. The amplified and equalized pulses are fed to a condenser *C* (the resistor $R_2$ serves for amplitude adjustment), whose voltage rises approximately in proportion to the number of pulses per second. This voltage is measured by the leakage current flowing through the adjustable part *R* of the resistor $R_1$ and is read off the milliammeter *mA* (cf. fig. 8), whose deflection can also be recorded by a separate strip chart recorder *B*.

The deflection of the milliammeter can be recorded continuously by a separate strip chart recorder (*fig. 12*), while the Geiger tube is moving along the goniometer scale. To effect the movement of the Geiger counter, a small synchronous motor is used. The chart in the recorder also being driven by a synchronous motor, a definite relationship is established between the abscissae of the chart and the Bragg angles, and the recorded curve immediately represents the intensity distribution in the diffraction pattern (*fig. 13*). The time for scanning 90 degrees can be varied between 6 hours and 45 minutes. For this purpose an interchangeable series of motors is provided, the one desired being connected to the vernier knob effecting the movement of the goniometer arm. (This is a simpler solution than a single stationary motor with controllable gear wheel drive.)

The intensity indicated by the milliammeter is a mean value, averaged over a time which depends on the product *RC* of the circuit described above

[6]) The first device of this kind, using thyratrons, was developed by C. E. Wynn-Williams: A thyratron "scale-of-two" automatic counter, Proc. Roy. Soc. A **136**, 312-324, 1932. A modern and widely adopted counting system is described by H. Lifschutz, A complete Geiger-Müller counter system, Rev. Sci. Instr. **10**, 21-26, 1939.

(fig. 11). A point of major importance in all work with Geiger counter tubes, and which will be considered in greater detail in the second article that is to appear subsequently, is the random time distribution of the X-ray quanta absorbed in the counter tube. The ensuing current intensity fluctuations would tend to render the position of a line peak indefinite if at a given scanning speed the averaging time (time constant) is made too short. On the other hand, if one adopts a very large time constant, it is evident that two closely adjacent diffraction lines will not be resolved because the averaging time will cover the time of traversing both lines. With a view to adapting the time constant to the scanning speed (and slit width) actually chosen, an adjustment by a suitable choice of the resistance $R$ is provided. For many purposes a preliminary rapid scanning of the spectrum, at a rate of e.g. 2 degrees/minute, will be desirable; a time constant of 2 seconds is suitable for this purpose. If subsequently the spectrum is more accurately scanned, e.g. at 0.25 degree/minute, the time constant may be set to say 5 seconds.

Finally, it may be mentioned that a test multivibrator is incorporated, delivering pulses at a fixed, known frequency and which, substituted for the Geiger counter tube, is used to check the functioning of both intensity measuring devices.

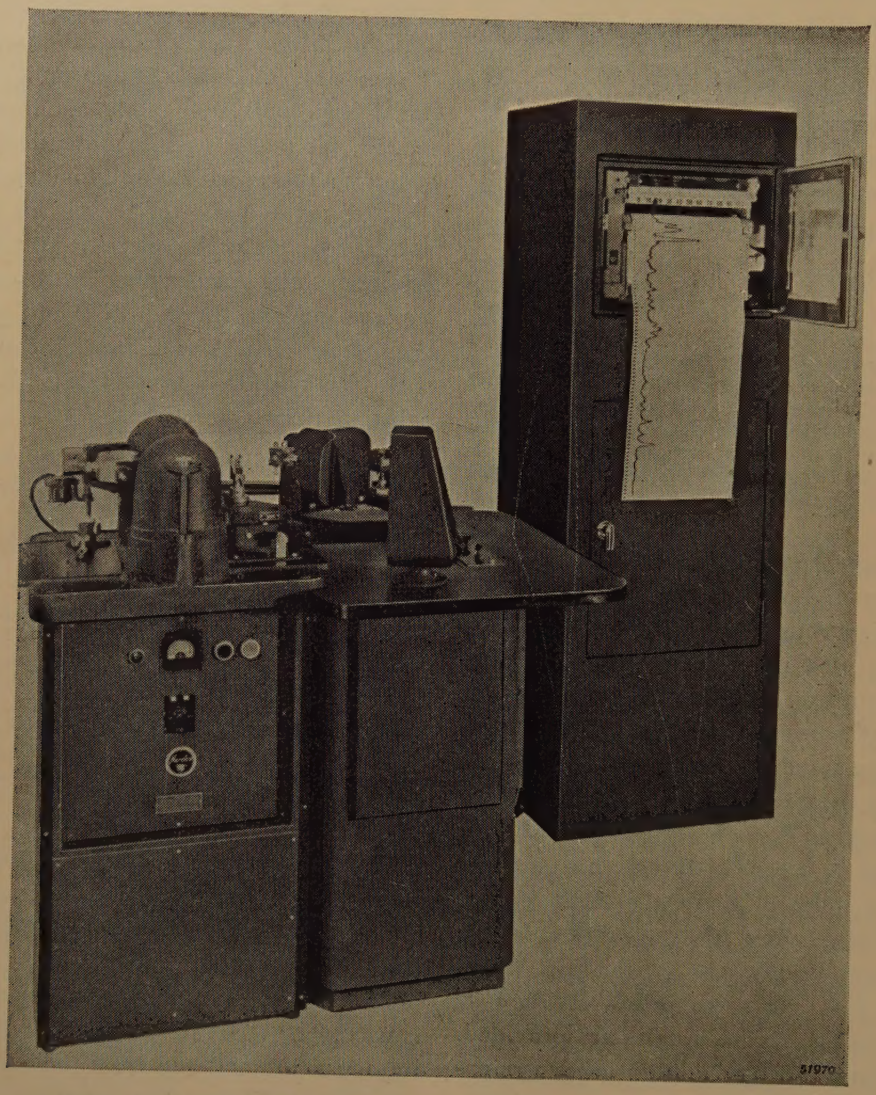


Fig. 12. The Geiger counter spectrometer with strip chart recorder.

## Adaptation of the Geiger counter tube to the diffraction measurements

To obtain the greatest possible sensitivity the Geiger counter ought to deliver a discharge pulse for every single X-ray quantum entering the receiving slit. With the design of the counter devel-

oped by H. Friedman [7]) this ideal is approached very closely. *Fig. 14*, representing a modified construction incorporated in most of the spectrometer units at present in use, shows the essential features of the design (cf. also fig. 5). The X-ray beam enters the tube at one end through a window of Lindemann glass, which is highly transparent for X-rays. This window, being 250 microns thick, transmits about 75% of the radiation with wavelength 1.5 Å (K$\alpha$-radiation of a copper target [8])). The beam then travels through the whole length of the tube more or less parallel to the central wire anode, at a distance of a few mm from this wire. It is essential for the beam to be absorbed as completely as possible in the gas filling of the counter tube, and not to strike the central wire nor the surrounding cylindrical cathode. A photo-electron is easily produced in the gas and almost certainly will trigger a discharge, while only a small number of photo-electrons would be liberated from the electrode metal and these would have a high probability of vanishing in the metal or elsewerhe (see below) without being counted.

The tube is in most cases filled with argon, which strongly absorbs the characteristic radiation of copper or iron targets (*fig. 15*). If a molybdenum target is used, a Geiger counter tube filled with krypton is better suited.

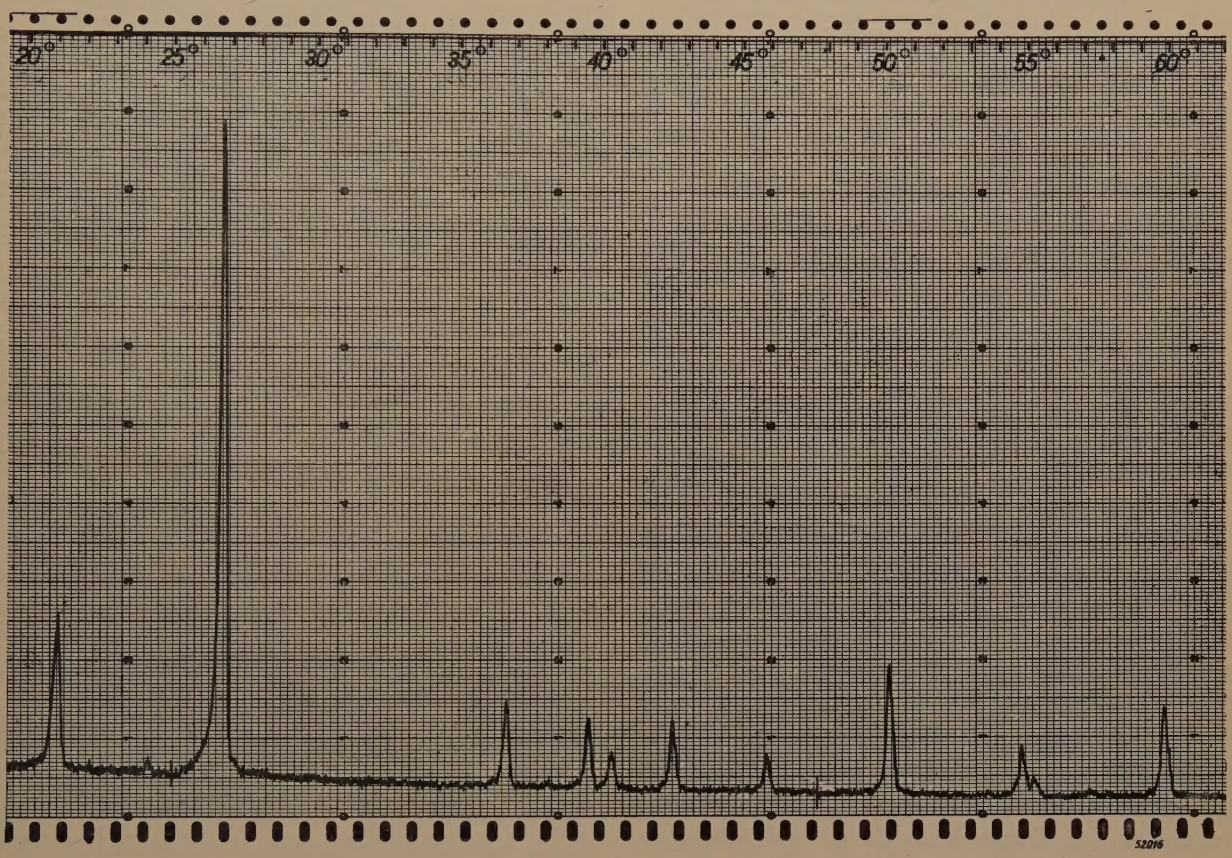



Fig. 13. Example of a diffraction pattern recorded with the Geiger counter spectrometer. The curve shows immediately the relative line intensities as a function of the Bragg angle ($2\Theta$). (The section of the spectrum shown here was recorded in appr. 3 hours.)

With an argon pressure of 30 cm Hg, 42% of the quanta of 1.5 Å wavelength entering the tube give rise to a discharge. Taking into account the absorption of the window, it is seen that an over-all efficiency of the counter of about 30% is achieved (percentage of quanta arriving at the tube window which is counted), a very high value indeed compared to the values of only a few percent obtained with former types of counter tubes.

During the past year still higher efficiencies have been attained by means of a counter tube provided with a mica window instead of the

[7]) H. Friedman, l.c. and U.S. Pat. 2386785.

[8]) The other components (K$\beta$-lines) of the copper radiation and part of the "white radiation", which produce a background and diffraction spectra of their own and whose superposition on the desired spectrum would cause serious errors, are eliminated by means of a thin nickel filter placed before the window.

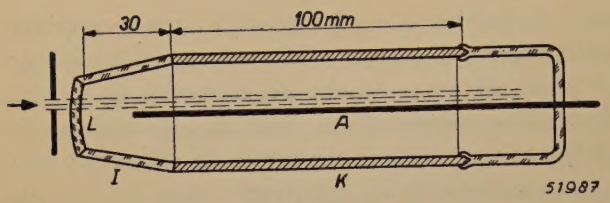



Fig. 14. Geiger counter tube for X-ray intensity measurements (longitudinal cross section of the tube shown in fig. 5). The X-ray beam enters the tube in a longitudinal direction and travels through it a short distance from and more or less parallel to the wire anode $A$. The entrance window $L$ is made of Lindemann glass, sealed to the chrome iron cathode cylinder $K$ via a short cylinder of soft glass $I$. The tube, if used for the measurement of copper or iron radiation, is filled with argon at a pressure of about 30 cm Hg and contains a small amount of an organic quenching vapor.

window of Lindemann glass. A tube of this type is shown in *fig. 16*. The mica window, having a thickness of only 12 microns, is sealed vacuum-tight to the chrome-iron cathode cylinder by a newly developed technique [9]). It transmits 86% of the incident copper $K\alpha$-radiation. Moreover, with this type of tube there is hardly any "dead space" between the window and the active counting volume. In the Lindemann window type of tube, the window could not be sealed directly to this metal cylinder; therefore, a short cylinder of soft glass was used as an intermediary (cf. figs 5 and 14). As the absorption of the X-rays in the gas is useful only when it takes place in the section within the cathode cylinder, the intermediate glass section entailed the existence of an inactive column of gas about 3 cm long, preceding the active column. Owing to the avoidance of this dead space in the mica window tubes, 55% of the entering quanta give rise to a discharge. Thus the over-all efficiency is increased to $0.86 \times 0.55 =$ about 50%, for the copper radiation.

The increase in efficiency is even more pronounced in the case of iron radiation, which is rather strongly absorbed in the Lindemann glass windows formerly used. Thus, measurements with iron radiation are considerably speeded up when using the mica window Geiger counter; still higher speed will be achieved when this counter tube is used in conjunction with an X-ray tube similarly provided with a mica window for the emerging X-rays.

As a consequence of the dead space in the Lindemann window tubes, the efficiency of these tubes exhibits an optimum at a definite gas pressure. Since the absorption by the gas increases with increasing gas pressure, a high pressure favors the complete absorption of the entering X-ray quanta in the tube — but then most of them are absorbed in the first, inactive part of the tube. With very low pressure, absorption in this part of the tube is negligible, but then the absorption in the useful part of the gas column also is far from complete and a considerable part of the quanta will pass through to the other end of the tube uncounted. Evidently, an optimum must exist at an intermediate pressure. The argon pressure actually chosen, about 30 cm Hg as mentioned above, is near the optimum.

With the mica window type of tube, in which practically no dead space with undesired absorption occurs, a higher argon pressure would be permissible, yielding a higher efficiency. However, a limit to the permissible increase of the pressure is set by the increase of the voltage which must be applied between the electrodes. We shall later return briefly to this point.

When the intensity of the X-ray beam is very high, the quanta enter the tube at such a rate that the time interval between some of them will be shorter than the "dead time" of the counter tube, i.e., the time necessary for the restoration of tube conditions necessary to ensure a response to a succeeding absorbed quantum. In this case not all

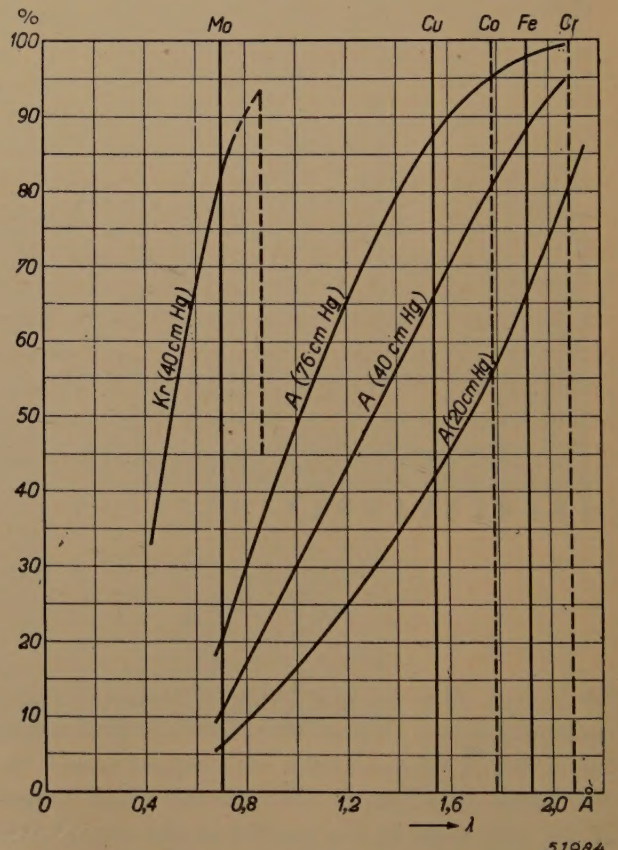



Fig. 15. Absorption of a column of 10 cm of argon (A) or krypton (Kr) for X-rays of different wavelengths, at different pressures. The characteristic radiation of copper (Cu $K\alpha$ = 1.5418 Å), is strongly absorbed by argon, as well as that of iron (1.9373 Å), chromium, cobalt. For the radiation of molybdenum (0.7107 Å), the argon filled Geiger tube is not effective and must be replaced by one filled with krypton. (The absorption edge, where the absorption suddenly becomes small with increasing wavelength, lies at 0.86 Å for krypton, at 3.87 Å for argon).

[9]) The technique is similar to the one described by J. S. Donal Jr. in Rev. Sci. Instr. **13**, 266, 1942; cf. also L. Malter, R. L. Jepsen and L. R. Blom, R. C. A. Review **7**, 623, 1946 (December).

absorbed quanta give rise to separate counts. Such counting losses, causing a non-linear response of the counting system to variations in X-ray intensity, affect the efficiency of the counter tube unfavorably and ultimately set a limit to the permissible intensity of the diffracted X-ray beam. In view of this undesirable limitation, measures are taken to make the dead time very short. A well-known and effective means is the addition of a small amount of an organic vapor to the gas. By a mechanism of a rather complex nature which is not yet completely understood, the vapor tends to reduce the number of charged particles available for maintaining the discharge so that it is rapidly quenched. Formerly, alcohol was generally used as a quenching component. This, however, had the drawback of being soon decomposed by the process in which it was involved. To-day other compounds are used, e.g., methylene bromide, which is more stable and equally efficient; thus, the life of a Geiger counter tube in constant use may amount to many months or even several years (the total number of counts performed amounting to the order of $10^9$), while the tubes lasted for at most a few months with the argon-alcohol filling. The quenching speed is such that up to 600 absorbed quanta per second may be counted without appreciable loss. However, counting is still possible at rates up to 3000 counts/second (taking into account the non-linearity correction). This corresponds to a beam intensity of approximately $10^4$ quanta per second arriving at the tube window (on the basis of a 30% overall counting efficiency). Since each quantum of the copper radiation at 1.5 Å represents an energy of approximately $10^{-8}$ erg, the permissible beam intensity is about $10^{-4}$ erg/sec. If necessary, the intensity can be kept within this limit by reducing the height of the primary beam divergence slit and the Geiger tube entrance slit, which is controllable between 1 and 6 mm by means of two horizontal wedge-shaped diaphragms.

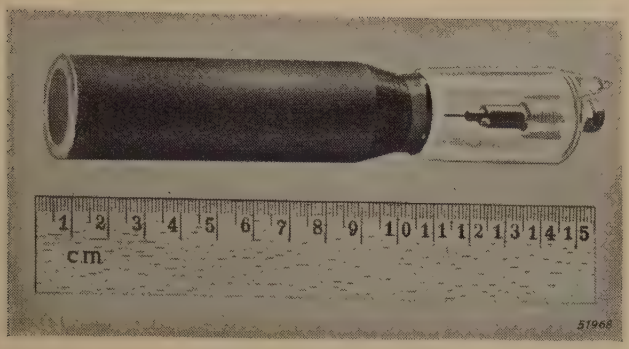



Fig. 16. Geiger counter tube, similar to that in figs 5 and 14, but provided with a mica window. The window is 12 microns thick and is sealed vacuum-tight to the chrome iron cathode cylinder. The axial anode wire, which nearly touches the window, carries a glass bead at its end, in view of the high field strength occurring at this point.

In connection with the last remark it is interesting to note that the intensity recorded by the counter tube does not vary linearly with the entrance slit height. This is due to the fact that in the outer part of the cylindrical gas column, near the wall, the field strength is much less than near the wire anode in the centre. The quenching vapor, trapping electrons as it is intended to do, eliminates many of the primary photo-electrons in the outer part of the tube which ought otherwise to produce a discharge pulse, while photo-electrons produced near the wire are accelerated rapidly enough to escape being passively absorbed. Thus there exists in the gas column a cylindrical "active volume", located around the anode wire. With the present type of counter tube at 3 mm from the wire the probability of an electron triggering a discharge is reduced to one half of the value at 1 mm.

We conclude this section with a few remarks on the voltage applied to the Geiger counter tube. This voltage should be chosen well up on the "plateau" of the counter tube, i.e. the voltage region where the desired intense short pulses with amplitude nearly independent of voltage can be obtained. Thus the functioning of the tube is made as nearly as possible independent of the voltage supply. On the other hand it is desirable to choose the voltage at the low voltage end of the "plateau", because during the life of the tube the plateau gradually shrinks at the high voltage end (and extends to lower voltages). This is due to the gradual disappearance of the quenching vapor. In our case a voltage of about 1400 V is applied, which at the argon pressure of 30 cm Hg (upper and lower limits of the plateau depend also on the gas pressure) is a suitable compromise.

### The use of the Geiger counter spectrometer

As the results obtainable with the spectrometer will be treated in a separate article, only a very brief survey is given here.

The peaks of diffraction lines can usually be located on spectrometer charts with an error no greater than $\pm$ 0.02°. This compares very favorably with the error of $\pm$ 0.1° generally inherent in photographic diffraction measurements. The resolving power likewise is beyond that attainable in the common photographic procedure. Usually the resolution of two diffraction lines obtained from one

lattice plane (spacing value $d$) with two slightly different wavelengths, e.g. Cu $K\alpha_1$ and $K\alpha_2$ (1.54050 and 1.54434 Å), is given as a criterion. From the Bragg equation it is easily seen that the angle difference for two such lines increases with angle $\Theta$ (diminishing $d$). On spectrometer recordings Cu $K\alpha_1$ and $K\alpha_2$ are clearly distinguishable even at $\Theta = 30°$ and completely resolved at $\Theta = 40°$, while on a film taken with a conventional powder camera (114 mm diameter) resolution of these lines starts only at $\Theta =$ about 50° and complete resolution is not obtained before $\Theta =$ about 60°.

One of the outstanding features of the new instrument is the ease with which quantitative measurements of relative line intensities are obtained. In the photographic procedure these measurements are rather difficult and time consuming because of their sensitivity to the exposure and developing conditions of the film and because of the necessary recording with a microphotometer. The advantage of the Geiger counter spectrometer becomes especially apparent if the position of the lines we are interested in is already known from a photographic diffraction pattern, whereupon the intensities of these lines can be accurately determined by scanning with the counter tube only those limited portions of the spectrum containing the lines under consideration. The time saving achieved in such cases can be quite spectacular.

A further advantage of the Geiger counter spectrometer as compared with the photographic method is the weaker and more evenly distributed background of the spectrum, which improves the possibility of measuring very weak lines and makes feasible the measurement of lines at very low Bragg angles (spacing values up to 150 Å). These points will be discussed more fully in the second article.

As to other specific applications of the spectrometer, it is evident that the use of a local radiation detector is ideally suited for experiments in which one or two diffraction lines must be continuously observed, e.g., the investigation of phase changes in a specimen. In such investigations it is also an important feature that, with the spectrometer, heating or cooling of the specimen is possible in a simple way without danger to the measuring device, and that, furthermore, the specimen is freely accessible for adjustments during the measurements. (Since of course the operator must not be exposed to the X-rays when making such adjustments, a shutter is incorporated which automatically cuts off the primary beam when the operator lifts the barrier flag located before the specimen.)

Finally, attention may be drawn to the fact that the charts obtained with the spectrometer offer no indication as to whether the intensity along a diffraction line is uniform or not. Therefore, special precautions are required in the preparation of the flat specimen in order to avoid a preferred orientation of the crystallites (texture) which would give rise to a non-uniform intensity distribution and thereby introduce errors into the intensity measurements.

# ELECTROMAGNETIC WAVES IN WAVE GUIDES

by W. OPECHOWSKI.

538.566.5:621.392.26

## PART I. GENERAL THEORETICAL PRINCIPLES; RECTANGULAR WAVE GUIDES.

In the domain of microwaves the ideas commonly held in regard to the theory of resonant circuits and transmission lines are limited in validity. A rigorous theory of the electromagnetic phenomena must therefore, in this field, be based directly upon the fundamental equations of the Maxwell theory. It is in this manner that the present article deals with the propagation of electromagnetic waves in what are called wave guides, *i.e.* long metal tubes as used in high-frequency technique for transmitting electromagnetic energy over short distances (*e.g.* inside an apparatus). In Part I the theory of rectangular wave guides is explained (in Part II, which will be published later, circular wave guides will be discussed). As an introduction the properties of plane electromagnetic waves in an unbounded conducting medium are dealt with and the methods for solving the basic equations of the Maxwell theory are summarized.

Resonant circuits employed in the range of decimeter and centimeter waves (microwaves) have been extensively discussed in a series of articles already published in this journal [1) 2) 3) 4)]. The last two articles particularly discussed the theoretical principles and practical applications of electromagnetic cavity resonators, dealing with the theory of cavity resonators mainly by means of a suitable generalization of concepts such as "capacitance", "self-inductance" and "impedance" with which we are familiar in the range of lower frequencies. We have said "mainly", for other problems also arose which could not be dealt with in this manner and in those cases it could only be stated what results are obtained by a direct solution of the Maxwell equations. As is known, these equations form the theoretical basis of all electromagnetic phenomena (excepting of course atomic structure of the charge and radiation as manifested in the existence of electrons and photons).

It is not surprising that when dealing with very high frequencies one is often at a loss to know what is to be understood by the concepts mentioned, for even the possibility of introducing these concepts, starting from the fundamentals of the Maxwell theory, arose from the very fact that one's interest was consciously limited to not too rapidly alternating electric currents and electromagnetic fields. Thanks to this limitation it became possible with the aid of these concepts to work out a theory of resonant circuits and transmission lines for this range. The great advantage of this theory is that it can be further applied with hardly any reference to the Maxwell equations, which are unnecessarily general for a description of the low frequency phenomena. That the equations of this theory are also for very high frequencies (let us say $\nu > 10^9$ c/s) in certain cases equivalent to the Maxwell equations is essentially due to the fact that the concept of electric (conduction) current holds for any frequency, since current is a quantity that occurs also in the Maxwell equations. In an imaginary strange world where exclusively high-frequency electromagnetic phenomena existed one would presumably not immediately have felt any call for introducing such concepts as capacitance and self-inductance.

The only generally valid manner of approaching the problems arising in the domain of very high frequencies is, therefore, to start directly from the basic equations of the Maxwell theory. By this we do not mean to say that it can never be desirable to follow some roundabout way that may be available via the range of lower frequencies with their specific concepts. On the contrary, firstly, because for anyone better acquainted with these concepts than with the Maxwell equations such can hardly constitute a roundabout method, and secondly because extension of the terminology of resonant circuits and transmission lines to the range of microwaves leads to a more uniform language, which in itself may be useful. As a matter of fact in this journal wide use will be made of the "language" of resonant circuits and transmission

---

1) C. G. A. von Lindern and G. de Vries, Resonant circuits for very high frequencies, Philips Techn. Rev. **6**, 217-224, 1941.
2) C. G. A. von Lindern and G. de Vries, Lecher systems, Philips Technical Rev. **6**, 240-249, 1941.
3) C. G. A. von Lindern and G. de Vries, Flat cavities as electrical resonators, Philips Techn. Rev. **8**, 149-160, 1946 (No. 5).
4) G. de Vries, Electromagnetic cavity resonators, Philips Techn. Rev. **9**, 73-84, 1947 (No. 3).

lines when discussing a device which for high frequencies performs a function of no less importance than the cavity resonator. We refer to the "wave guides", serving to convey electromagnetic energy within a transmitting or receiving apparatus from one point to another. Wave guides are long metal tubes which may be, for instance, rectangular, circular or elliptical in cross section and through which electromagnetic waves can be propagated. Governing the whole of the high-frequency technique is the phenomenon that owing to skin effect the current flows only in a thin surface layer of a conductor and scarcely does anything more than dissipate energy. Consequently the energy transmission takes place substantially not in the conductor but by means of an electromagnetic wave around the conductor or, as in the case of wave guides, inside a space bounded by conductors. The function of the conductors is merely to mark out, as it were, the path for the wave.

In this article we shall study the propagation of electromagnetic waves in cases where conductors are present (particularly in wave guides), starting from the basic equations of the Maxwell theory. We shall only incidentally consider in how far the same results could be deduced by an extension of the concepts commonly used for low frequencies; the question where the limits of such an extension lie has been sufficiently dealt with in the articles already quoted [5]). Neither shall we enter into the engineering aspect of wave guides, this being reserved for other articles to be subsequently published in this journal. Of course we shall not omit to bring forward whatever may be of particular technical importance. In the following section a summary is given of the general mathematical formulation of the fundamentals of the Maxwell theory. The actual subject of the present article begins with the next section on plane waves.

## Basic equations of the Maxwell theory

In the Maxwell theory the electromagnetic phenomena are described with the aid of two kinds of quantities, one relating to the electric charges at rest and in motion, the other relating to the electromagnetic field.

As a rule two quantities suffice to describe the behaviour of the electric charges: the charge density $\varrho$, *i.e.* the amount of charge per unit volume, and the current density **J**, *i.e.* the amount of charge which per unit time passes a unit surface perpendicular to the direction of motion; by printing **J** in heavy type we indicate that it is a vector. The quantities $\varrho$ and **J** are not independent of each other, for when we consider a volume $\tau$ containing electric charges and enveloped by a closed surface $S$ then, in virtue of the law of conservation of the charge, the following equation must hold:

$$\oint\!\!\oint J_n \, dS = -\frac{d}{dt} \iiint \varrho \, d\tau, \quad \ldots\ldots \quad (I)$$

where $t$ is the time and $J_n$ the component of **J** along the normal to the surface at the point considered on the surface; as positive direction of the normal the direction outwards is chosen (we shall consistently keep to this latter convention in this article). The integral on the left means that the surface $S$ has to be divided into infinitesimal elements $dS$, the product $J_n \cdot dS$ then having to be found for each element and all these products added up. The small circles remind us that $S$ is a closed surface.

The electromagnetic field is usually characterized by four vectors whose physical significance is here assumed to be known: the electric field strength **E**, the magnetic field strength **H**, the magnetic induction **B** and the electric displacement **D**. The essence of the Maxwell theory is now comprised in two equations correlating **E** and **B** respectively **H** and **D**.

One equation reads as follows:

$$\oint E_l \, dl = -\frac{d}{dt} \iint B_n \, dS \ldots\ldots\ldots \quad (II)$$

Here the integral on the right extends over an arbitrary surface $S$ bounded by the closed curve (contour) $l$ along which the integral on the left has been taken; the curve $l$ is as a matter of fact an arbitrary curve. $B_n$ is the component of **B** at the point considered on the surface $S$ in a direction perpendicular to the surface. The meaning of the integral on the right is analogous to that of the integral of the left-hand side of eq. (I) except that there the surface was closed. $E_l$ is the component of **E** at the point considered on the curve $l$ along the tangent at that point. The meaning of the "line integral" of the left-hand side of eq. (II) — where the circle indicates that $l$ is a closed curve — is roughly speaking as follows: the curve $l$ is divided into infinitesimal elements $dl$; the product $E_l \cdot dl$ is found for each element and all these products are added up. The minus sign in eq. (II) is necessary to conform to what has been agreed upon concerning the positive directions of the normal to the surface and of the tangent to the curve; in this article we shall follow consistently the "corkscrew rule", well known from the text books (*fig. 1*).

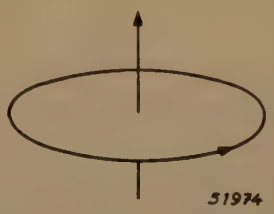



Fig. 1. The relation between the positive direction along a contour and the positive direction of the normal of the surface bounded by this contour, according to the usual convention ("corkscrew rule").

Eq. (II) gives expression to Faraday's law of induction. If we imagine the curve $l$ to be formed by a wire then the integral on the left-hand side of the equation represents the e.m.f. induced in the wire by the variation of the magnetic flux $\iint B_n dS$; the validity of eq. (II) however is postulated in the Maxwell theory also for vacuum, that is to say also in the

[5]) See particularly pp. 222-223 of the article quoted in footnote 1 and pp. 78-79 of the article quoted in footnote 4.

absence of any conductor that makes the contour $l$ as it were tangible. If, as is usual in electrotechnics, the e.m.f. is to be expressed in volts and we agree to measure length in metres and time in seconds, then from eq. (II) it immediately follows that the unit of **E** is volt/m and that of **B** volt · sec/m².

The other fundamental equation of the Maxwell theory is:

$$\oint H_l \, dl = \iint J_n \, dS + \frac{d}{dt} \iint D_n \, dS \quad . \quad . \quad . \quad \text{(III)}$$

Here again on the right-hand side we have an integration over an arbitrary surface bounded by an also arbitrary contour along which the integral on the left has been taken. The meaning of the symbols is analogous to that of the symbols in eq. (II). Supposing for a moment that **D** is independent of time, then the second term on the right in eq. (III) is zero. The equation then gives expression to the well-known fact that the "magnetomotive force", *i.e.* the line integral of the left-hand side measured along a curve enclosing a conductor (*e.g.* a wire), is equal to the total current intensity flowing in the conductor. The second term on the right in eq. (III) represents the so-called displacement current introduced by Maxwell in his theory. The necessity of introducing such a term can be illustrated with all kinds of examples, which we shall not enter upon here. We would observe however that if the displacement current were to be omitted from equation (III) we should be acting contrary to the law of the conservation of charge (eq. (I)).

Expressing the total current intensity in amperes — again in accordance with technical usage — and measuring length in metres and time in seconds in the same way as before, it follows from eq. (III) that the unit of **H** is amp/m and that of **D** amp · sec/m².

From equations (I), (II) and (III) it can further be immediately deduced that:

$$\oint\oint B_n \, dS = \text{constant},$$

$$\oint\oint D_n \, dS - \iiint \varrho \, d\tau = \text{constant}.$$

The physical significance of these equations leads to the conclusion that the constants on the right-hand side of the two equations have to be put equal to zero (think, in particular, of magnetostatics and electrostatics!). Hence:

$$\oint\oint B_n \, dS = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(IV)}$$

$$\oint\oint D_n \, dS = \iiint \varrho \, d\tau \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(V)}$$

Equations (IV) and (V) thus fulfil the function of supplementary conditions which **B** and **D** have to satisfy.

The two basic equations (II) and (III) of the Maxwell theory are absolutely general, in the sense that nothing is assumed in regard to the electromagnetic properties of the medium. They are not, however, sufficient to determine the five quantities **E**, **B**, **H**, **D** and **J**. They only give a relation between **E** and **B** (eq. (II)) and between **H**, **D** and **J** (eq. (III)) but say nothing about the relation of **E** and **B** on the one hand to **H**, **D** and **J** on the other hand. It is in the nature of the latter relation that the electromagnetic properties of the medium find expression, that is to say, as a rule this relation differs according to the medium in which the electromagnetic field may be present. For homogeneous, isotropic media the following relations mostly [6]) hold with sufficient approximation:

$$\mathbf{D} = \varepsilon \mathbf{E}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(VI)}$$

$$\mathbf{B} = \mu \mathbf{H}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(VII)}$$

$$\mathbf{J} = \sigma \mathbf{E}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(VIII)}$$

in which the dielectric constant $\varepsilon$, the magnetic permeability $\mu$ and the specific conductivity $\sigma$ are constants depending upon the nature of the medium. The units in which these three constants are expressed follow immediately from the units chosen above for **D**, **E**, **B**, **H** and **J**. The unit of $\varepsilon$ is amp·sec·volt$^{-1}$·m$^{-1}$ or (definition of coulomb) coulomb·volt$^{-1}$·m$^{-1}$ or (definition of farad) farad·m$^{-1}$, that of $\mu$ is volt·amp$^{-1}$·sec·m$^{-1}$ or (definition of henry) henry·m$^{-1}$ and that of $\sigma$ is amp·volt$^{-1}$·m$^{-1}$ or (definition of ohm) ohm$^{-1}$·m$^{-1}$.

Substituting in the equations (II) and (III) the expressions (VI, VII, VIII) for the quantities **D**, **B** and **J**, we obtain two equations between the quantities **E** and **H** (eqs. (1) and (2) below) which will form the basis of what follows in this article. By performing the same substitution in (IV) and (V) we obtain two supplementary equations (eqs. (3) and (4) below).

In the foregoing we have used the terms volt and ampere without any further definition; the other electromagnetic units have been defined with the aid of volt, ampere, metre (as length unit) and second (as time unit). By expressing all quantities in volts, amperes and the units derived therefrom in the manner indicated above we have introduced the so-called rationalized Giorgi system of electromagnetic units. The Giorgi system will be discussed separately in two articles that will appear shortly in this journal, when the definition of volt and ampere will be given. Here we shall only mention that the magnetic permeability of vacuum in this system has the value:

$$\mu_0 = \frac{4\pi}{10^7} \frac{\text{henry}}{\text{metre}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(IX)}$$

From (IX) and a relation which we shall deduce later, *viz.* eq. (16), it follows that the value of the dielectric constant $\varepsilon_0$ of vacuum is:

$$\varepsilon_0 = \frac{10^7}{4\pi c^2} \frac{\text{farad}}{\text{meter}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(X)}$$

in which $c$ indicates the numerical value of the velocity of light in m/sec.

## Plane waves in an unbounded conducting medium

We shall first apply the fundamental equations of the Maxwell theory to the case of an unbounded, homogeneous and isotropic medium, which is thus characterized by given, constant values of $\varepsilon$, $\mu$ and $\sigma$. (These constants may still be functions of the frequency of the electromagnetic waves considered.)

In fact such a case does not occur in practice, but it will serve as starting point for the investigation of wave guides. Moreover, already in this case we shall obtain certain results not without interest for the interpretation of technically important facts. The equations for the electric and magnetic field strengths **E** and **H** then assume the following form:

$$\oint E_l \, dl = -\mu \frac{d}{dt} \iint H_n \, dS, \quad . \quad . \quad . \quad . \quad \text{(1)}$$

[6]) We say "mostly" because for ferromagnetic media (particularly iron) equation (VII) no longer applies, and for substances showing the Hall effect (VIII) is no longer correct.

$$\oint H_l \, dl = \sigma \iint E_n \, dS + \varepsilon \frac{d}{dt} \iint E_n \, dS ; \quad . \quad (2)$$

to which are to be added the two conditions which the solutions of (1) and (2) have to satisfy:

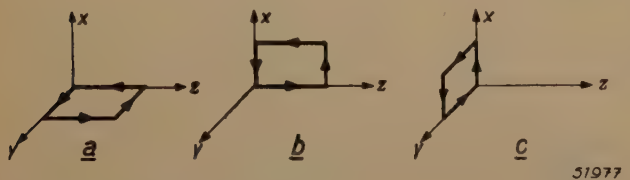
$$\oint \oint H_n \, dS = 0, \quad . \quad . \quad . \quad . \quad . \quad (3)$$

$$\varepsilon \oint \oint E_n \, dS = \iiint \varrho \, d\tau . \quad . \quad . \quad (4)$$

Here it is assumed that all quantities are expressed in units of the rationalized Giorgi sytems. The notation employed has been explained in the preceding section.

By choosing suitable integration contours and surfaces we shall be able to deduce differential equations valid for such a medium. We shall confine ourselves to the particular case where **E** and **H** depend only upon one coordinate, for instance $z$, of a rectangular system of coordinates, and the time. This suffices for our purpose, for what we want to show is that a plane wave propagated in any direction is a solution of the fundamental equations (1) and (2), after which we shall investigate the properties of such a wave; here the word "plane" means that in any plane perpendicular to the direction of propagation the **E** and **H** vectors respectively are everywhere equal at any given moment; if we now take the $z$-axis parallel to the direction of propagation then **E** and **H** are indeed dependent only upon $z$ and $t$.

In order to deduce the said differential equations we first apply eq. (1) to the contour shown in *fig. 2a*; as surface $S$ we take the rectangle bounded




51977

Fig. 2. Integration contours for deduction of equations (5)-(6) from equations (1)-(2).

by this contour in the $y$-$z$ plane. Since the contributions to the line integral of the two lengths of line parallel to the $z$-axis cancel each other (**E** and **H** depending only upon $z$ and $t$) we find

$$y \cdot E_y(0,t) - y \cdot E_y(z,t) = \mu y \int_0^z \frac{\partial}{\partial t} H_x(\zeta, t) \, d\zeta .$$

By differentiation with respect to $z$ we obtain

$$\frac{\partial E_y}{\partial z} = \mu \frac{\partial H_x}{\partial t} . \quad . \quad . \quad . \quad . \quad . \quad (5a)$$

By applying eq. (1) to the contours of figs 2*b* and 2*c* we find in the same way:

$$-\frac{\partial E_x}{\partial z} = \mu \frac{\partial H_y}{\partial t} . \quad . \quad . \quad . \quad . \quad . \quad (5b)$$

and

$$0 = \mu \frac{\partial H_z}{\partial t} . \quad . \quad . \quad . \quad . \quad . \quad (5c)$$

By carrying out the same operations with eq. (2) we obtain:

$$-\frac{\partial H_y}{\partial z} = \sigma E_x + \varepsilon \frac{\partial E_x}{\partial t}, \quad . \quad . \quad . \quad . \quad (6a)$$

$$\frac{\partial H_x}{\partial z} = \sigma E_y + \varepsilon \frac{\partial E_y}{\partial t}, \quad . \quad . \quad . \quad (6b)$$

$$0 = \sigma E_z + \varepsilon \frac{\partial E_z}{\partial t} . \quad . \quad . \quad . \quad . \quad (6c)$$

Thus six simultaneous differential equations (5)-(6) have been derived for the six quantities $E_x \ldots H_z$, each of which is a function of only $z$ and $t$.

We shall now prove that a plane wave running in the $z$ direction represents a solution of (5)-(6).

For the sake of simplicity, in what follows we shall consider only linearly polarized, harmonic plane waves.

"Linearly polarized" means that **E**, respectively **H**, has always the same direction for any value of $z$. By "harmonic" is to be understood that **E** and **H** depend sinusoidally upon $z$ and $t$. The mathematical expression for such a plane wave is, therefore, in the well-known complex representation:

$$\left.\begin{aligned} \mathbf{E} &= \mathbf{E}^\circ \, e^{j(\omega t - kz + \eta)}, \\ \mathbf{H} &= \mathbf{H}^\circ \, e^{j(\omega t - kz)}, \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (7)$$

where the symbols have the following meanings: $\mathbf{E}^\circ$, $\mathbf{H}^\circ$ are two real constant vectors (of which we know nothing more at the moment); the length of the vectors $\mathbf{E}^\circ$ and $\mathbf{H}^\circ$ will be denoted by $E^\circ$ and $H^\circ$ respectively;

$\omega = 2\pi\nu$ radial frequency, ($\nu$ frequency),
$\eta$ phase difference between the **E** and **H** vectors,
$j$ the imaginary unit, $+\sqrt{-1}$,
$k$ a constant.

It will appear that (7) is indeed a solution of (5)-(6), provided $k$, $\eta$ and $E^\circ/H^\circ$ depend in a certain manner upon $\mu$, $\varepsilon$, $\sigma$ and $\omega$. This dependence is found by substituting (7) in eq. (5)-(6) and

seeking the consequences of the requirement that (7) is a solution of these equations.

With a suitable choice of the system of coordinates this substitution yields:

$$\left.\begin{aligned} kE^\circ &= \mu\,\omega\,H^\circ\,e^{-j\eta}, \\ (\sigma + j\omega\varepsilon)\,E^\circ &= j\,k\,H^\circ\,e^{-j\eta}, \end{aligned}\right\} \quad . \; . \; . \quad (8)$$

where

$$E^\circ = E_x^\circ,\; H^\circ = H_y^\circ,\; E_y^\circ = H_x^\circ = 0, \quad (9a)$$

$$E_z^\circ = H_z^\circ = 0 \; . \; . \; . \; . \; . \quad (9b)$$

Eq. (9b) expresses the fact that the wave is transverse, whereas equations (9a) and (9b) together express the fact that the electric vector and the magnetic vector are at right-angles to each other.

Further from (8) it immediately follows that:

$$k^2 = \mu\omega\,(\varepsilon\omega - j\sigma), \; . \; . \; . \; . \quad (10)$$

$$\frac{E^\circ}{H^\circ} = \frac{\mu\omega}{k\,e^{j\eta}}.$$

Since we have assumed that $E^\circ$, $H^\circ$, $\mu$ and $\omega$ are real, according to the last equations also $k\,e^{j\eta}$ must be real; $k$ is therefore a complex number for which the following holds:

$$k = |k|\,e^{-j\eta}, \; . \; . \; . \; . \; . \; . \quad (11)$$

so that

$$\frac{E^\circ}{H^\circ} = \frac{\mu\omega}{|k|} \; . \; . \; . \; . \; . \; . \; . \quad (12)$$

According to (12) the ratio $E^\circ/H^\circ$ is always positive. Since the chosen system of coordinates $x$, $y$ and $z$ is a right-handed system (*cf.* fig. 2) this means that the directions $\mathbf{E}^\circ$, $\mathbf{H}^\circ$ and the positive $z$-direction also form a right-handed system.

Thus we have found a solution of (5)-(6) which, apart from the sign of $k$ and an arbitrary factor by which it can be multiplied, is determined unambiguously by (7) and (10)-(12).

Strictly speaking we should also verify that this solution complies with the differential conditions following from (3)-(4) just as (5)-(6) follow from (1)-(2). Such is not difficult to prove if need be. Suffice it to state that it is thereby found that the charge density $\varrho$ must everywhere be zero.

From (10) we see that in general the constant $k$ is a complex number, namely in all cases where $\sigma \neq 0$, that is to say when the medium is conducting. We shall see presently what this means. First we shall consider the case where $\sigma = 0$, i.e. where the medium is a perfect insulator. Then, according to the eq. (10)-(12):

$$k = \omega\sqrt{\mu\varepsilon}, \; . \; . \; . \; . \; . \; . \quad (13)$$

$$E^\circ/H^\circ = \sqrt{\mu/\varepsilon}, \; . \; . \; . \; . \; . \quad (14)$$

$$\eta = 0,$$

the last equation signifying that $\mathbf{E}$ and $\mathbf{H}$ vibrate in phase. The physical significance of $k$ is clear when bearing in mind that the wavelength $\lambda$ of the wave (7) in the case in question ($k$ real) is given by

$$\lambda = 2\pi/k \, . \; . \; . \; . \; . \; . \; . \quad (15)$$

From the well-known relation between the phase velocity $v$ of a wave and its frequency and wavelength $v = \lambda\nu$ it therefore follows that

$$v = \frac{1}{\sqrt{\mu\varepsilon}}.$$

If the medium concerned is a vacuum then this phase velocity is equal to $c \approx 3 \cdot 10^8$ m/sec, the velocity of light in vacum, that is to say:

$$c = \frac{1}{\sqrt{\mu_0\varepsilon_0}}\ \text{m/sec} \; . \; . \; . \; . \; . \quad (16)$$

Now (see eq. (IX))

$$\mu_0 = 4\pi/10^7\ \text{volt·sec/amp·metre}.$$

For the relation of amplitudes $E^\circ/H^\circ$ we thus get in this case

$$\frac{E^\circ}{H^\circ} = \sqrt{\frac{\mu_0}{\varepsilon_0}} = \frac{4\pi c}{10^7} \approx 377\ \text{ohms} \; . \quad (17)$$

Moreover we find:

$$k = \frac{\omega}{c}.$$

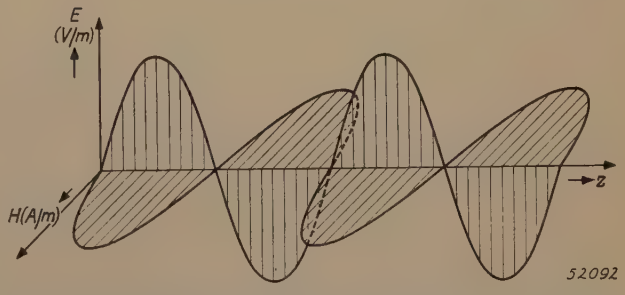



Fig. 3. Diagrammatic representation of the form of a linearly polarized, sinusoidal plane wave in vacuum ("instantaneous view").

The form of the wave in vacuum is represented diagrammatically in *fig. 3*.

If the medium is conducting ($\sigma > 0$), then $k$ is complex and may be written as

$$k = k_1 - jk_2, \; . \; . \; . \; . \; . \quad (18)$$

where $k_1$ and $k_2$ are real numbers. On account of (11) we therefore have

$$\tan \eta = k_2/k_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

Substituting (18) in (7) we obtain

$$\left.\begin{aligned} \mathbf{E} &= \mathbf{E}^\circ \, e^{-k_2 z} \, e^{j(\omega t - k_1 z + \eta)}, \\ \mathbf{H} &= \mathbf{H}^\circ \, e^{-k_2 z} \, e^{j(\omega t - k_1 z)}. \end{aligned}\right\} \quad . \quad . \quad . \quad (20)$$

These formulae represent a wave damped in the direction of propagation, provided $k_1$ and $k_2$ are positive. Now $k_1$ and $k_2$ are determined by (10), except only for the sign. Let us therefore assume that $k_2$ for instance is always positive. It appears that then also $k_1$ is always positive. Thus we see that a plane electromagnetic wave in a conducting medium must necessarily be damped, or, in other words, there is then always an absorption of the wave. This is not surprising, for, if $\sigma > 0$, then the wave induces a (conduction) alternating current (density $\sigma\mathbf{E}$) in the medium; the Joule heat of this current is generated at the cost of the energy of the wave, this being manifested in the decrease of the wave amplitudes. Now the Joule heat amounts to $\sigma E^2$ joule/m$^3$. If the medium is a very good conductor (large $\sigma$) then consequently the damping is strong, so that a wave penetrating such a medium travels only a very short distance before it becomes no longer perceptible. The greater the value of $\omega$, the shorter is this distance. In a perfect conductor ($\sigma = \infty$) wave propagation is absolutely impossible. As a matter of fact, from the fundamental equations (1)-(2) it is immediately seen that in a perfect conductor the **E**, **H** alternating field must always be zero. We would further point out that in a conducting medium **E** and **H** no longer oscillate in phase ($\eta \neq 0$) and that both the wavelength and the ratio $E^\circ/H^\circ$ are smaller than in the case of vacuum. In the limiting case of a very good conductor the form of the plane wave undergoes such an alteration in consequence of the strong damping that the sinusoidal character of the wave can hardly be recognized, as may be seen from *fig. 4*.

Considering the form of eq. (10) there are obviously two limiting cases to be discussed separately, namely $\sigma/\varepsilon\omega \ll 1$ and $\varepsilon\omega/\sigma \ll 1$. Since the total current density is given by

$$\sigma \mathbf{E} + \varepsilon \frac{\partial \mathbf{E}}{\partial t} = (\sigma + j\omega\varepsilon)\,\mathbf{E},$$

it corresponds in the first case to a practically pure displacement current: the medium behaves as an insulator. In the second case the total current is practically a conduction current, the medium acting as a good conductor.

*First limiting case:* $\sigma/\varepsilon\omega \ll 1$.

In this case the absorption of the electromagnetic wave is often called the dielectric absorption and one speaks also of dielectric losses. One then sometimes introduces a complex dielectric constant, $\underline{\varepsilon}$, that is to say one writes (10) in the form:

$$k^2 = \omega^2 \mu \underline{\varepsilon}$$

where

$$\underline{\varepsilon} = \varepsilon - j\sigma/\omega.$$

The well-known "loss angle" $\delta$, which may be defined as the angle whose tangent is equal to the absolute value of the ratio of the imaginary to the real part of $\underline{\varepsilon}$, stands in a very simple relation to the phase difference $\eta$ between **E** and **H**, for from the last two equations and eq. (11) it follows that

$$\delta = 2\eta, \qquad \tan \delta = \tan 2\eta = \sigma/\varepsilon\omega.$$

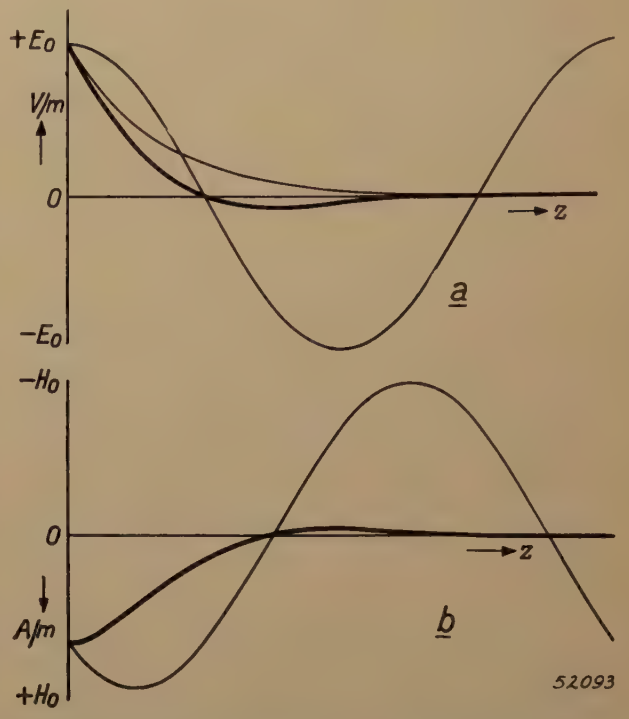



Fig. 4. Diagrammatic representation ("instantaneous view") of the structure of a linearly polarized, sinusoidal, damped plane wave in a well-conducting medium at high frequency ($\varepsilon\omega/\sigma \ll 1$). In (*a*) the value of the electric field strength and in (*b*) the value of the magnetic field strength are plotted vertically as functions of the distance $z$ along the direction of propagation. The lightly-drawn sine curves represent the wave when there is no damping. By multiplying the values of these sine functions by the corresponding values of the damping factor (see (*a*), lightly drawn) one obtains the heavy curves, which thus represent the damped wave. There is a difference in the shape of these curves in (*a*) and (*b*) due to the phase difference between **E** and **H**, which amounts here to $\pi/4$. Since **E**, **H** and the direction of propagation are at right angles to each other and, moreover, form a right-handed system, in order to obtain a three-dimensional picture one would have to turn the plane of the drawing in (*b*) 90° about the $z$-axis, such that the positive direction of the $H$-axis points towards the reader. Both the ratio $E^\circ/H^\circ$ and the wavelength are a factor of the order of $\sqrt{\varepsilon\omega/\sigma}$ smaller for a conducting medium than for vacuum.

The loss angle is therefore equal to twice the phase difference between **E** and **H**. From eq. (20) we see that the wavelength of the wave in a dielectric is $2\pi/k_1$. The damping of the wave amplitudes per wavelength is therefore, from (19):

$$e^{-k_2\lambda} = e^{-2\pi k_2/k_1} = e^{-2\pi\tan\eta} = e^{-2\pi\tan(\delta/2)}.$$

All these formulae are still exact. Introducing the assumption that $\sigma/\varepsilon\omega \ll 1$, we find with the aid of (10) and (18)

$$k_1 \approx \omega\sqrt{\mu\varepsilon}, \quad k_2 \approx \frac{\sigma}{2}\sqrt{\frac{\mu}{2}}, \quad \tan\eta \approx \eta = \frac{\delta}{2} = \frac{\sigma}{2\omega\varepsilon}.$$

A comparison of these formulae with (13)-(15) shows that to a first approximation the wavelength and the ratio $E^\circ/H^\circ$ have here the same values as those for a wave in a non-conducting medium. Little is noticed of the damping of the wave amplitudes so long as $\sigma/\varepsilon\omega$ is very small, the damping factor per wavelength being $e^{-k_2\lambda} \approx e^{-\pi(\sigma/\varepsilon\omega)}$.

*Second limiting case:* $\varepsilon\omega/\sigma \ll 1$.

One may then to a first approximation entirely ignore the displacement current, that is to say the first term in the right-hand member of (10) can be omitted. We then obtain:

$$k^2 \approx -j\mu\sigma\omega$$

and since

$$\sqrt{-j} = (1-j)/\sqrt{2},$$

$$k \approx \sqrt{\frac{\mu\omega\sigma}{2}}\,(1-j),$$

that is to say

$$k_1 \approx k_2 \approx \sqrt{\frac{\mu\omega\sigma}{2}}, \quad \eta \approx \frac{\pi}{4}, \quad \delta \approx \frac{\pi}{2}.$$

A comparison of these formulae with (13)-(15) shows that in a good conductor an electromagnetic wave assumes an entirely different form from that in a non-conducting medium (see fig. 4). Here the ratio $E^\circ/H^\circ$ and the wavelength are smaller than in a non-conducting medium by factors $\sqrt{\varepsilon\omega/\sigma}$ and $\sqrt{2\varepsilon\omega/\sigma}$ respectively (with $\varepsilon\omega/\sigma \ll 1$!). Moreover the damping is so strong that over a distance of one wavelength the field amplitudes become smaller by a factor 0.002, for in this case the damping factor of a wavelength amounts to

$$e^{-k_2\lambda} = e^{-2\pi k_2/k_1} \approx e^{-2\pi} \approx 0.002.$$

This strong absorption of the electromagnetic wave is closely related to the skin effect. From the theory of skin effect it is known that the so-called penetration depth $d$ is given by

$$d = \sqrt{\frac{2}{\mu\omega\sigma}}$$

and this is found to be equal to $1/k_2$.

The significance of the penetration depth $d$ follows from the fact that the amplitude of density of a high-frequency current in a conductor decreases proportionately with $e^{-z/d}$, where $z$ represents the depth below the surface of the conductor. Now, in consequence of (20), the amplitudes of the E and H waves diminish according to exactly the same law with the same value of $d = 1/k_2$. This correspondence is easy to understand when the high-frequency current is construed as being caused by the electromagnetic field tending to penetrate into the conductor from the outside.

If for instance $\lambda = 10$ cm (in vacuum), then from the formula given above for the penetration depth $d$ we find that for copper ($\sigma = 5{\cdot}8 \times 10^7$ ohm$^{-1}$ $\times$ m$^{-1}$) this amounts to only $1{\cdot}2 \times 10^{-6}$ m.

## Propagation of the waves in a homogeneous medium bounded by conductors

### *Boundary conditions*

Contrary to what has been assumed in the preceding section, in reality a homogeneous medium is always bounded. The homogeneous medium air, for instance, in a cavity resonator or in a wave guide is bounded by metal walls. Such walls can of course be regarded as another homogeneous medium. The values of $\varepsilon$, $\mu$ and $\sigma$ therefore make a jump on the boundary surface between air and metal. A similar situation occurs on the boundary between any two homogeneous media.

From what has been stated so far about the mathematical principles of the Maxwell theory it is by no means clear how such a situation has to be dealt with mathematically. We do know that plane (damped) waves can be propagated in each of the two media, but we do not yet know how the amplitudes and phases of a plane wave in one medium are to be "matched" with those of a plane wave in the other medium on the boundary surface. The fact that this matching cannot be arbitrary, so that some "matching rule" is essential, follows at once from the existence of the empirically established laws for the passage of a plane wave from one medium to another (laws of refraction and reflection).

It appears that such a general (*i.e.* not limited to plane waves) "matching rule", which finds expression in the existence of certain boundary conditions that have to be satisfied by **H**, **B**, **E** and **D** on the boundary surface, is indeed to be deduced from Maxwell's fundamental equations. Here we shall not formulate this matching rule in a general way but only for the particular case in which one of the two media is a perfect conductor ($\sigma = \infty$). Throughout this article metals will be replaced by perfect conductors in which, therefore, the electromagnetic field is exactly zero. This is certainly permissible (and in fact usual) as a first approximation, since, as we have seen in the example of plane waves, when the frequency is high the electromagnetic field scarcely penetrates into a good conductor. If the energy losses are of any interest — and in practice that is of course the case — then in a further approximation corrections have to be made for this slight penetration of the field into the metal.

The reason why the conductors are mostly assumed to have an infinite conductivity is that the boundary conditions become very simple if the electromagnetic field inside one of the two media is exactly zero. In the case that both media have finite conductivities the boundary conditions lead, on the other hand, in most problems to great mathematical difficulties.

The boundary conditions for a perfect conductor may now be formulated as follows. One resolves

E on the boundary surface between the conductor and the other medium (not necessarily vacuum) into a tangential component $E_t$ and a normal component $E_n$; as positive direction of the normal we take the direction from the conductor to the other medium. The same is done with H. We then have on the boundary surface:

$$E_t = 0, \quad H_n = 0, \quad . \quad . \quad . \quad . \quad . \quad (21)$$

$$\varepsilon E_n = s, \quad H_t = i, \quad . \quad . \quad . \quad . \quad . \quad (22)$$

where $s$ and $i$ stand respectively for the density of a surface charge (in coulomb/m$^2$) and the density of a surface current (in ampere/m); this current is at right-angles to the direction of H, such that the (positive) direction of the normal of the surface, the direction of H and the direction in which $i$ flows form a right-handed system.

In order to understand better the significance of the "surface current", think for instance of a wire through which a high frequency alternating current is flowing. Owing to the skin effect the density of this current will have measurable values only in a thin surface layer of the wire. Roughly speaking one may therefore say that the current flows "along the surface" of the wire. The density of this "surface current" will then be equal to the current intensity (in amperes) divided by the circumference (in metres) of the cross section of the wire. The density calculated in this manner is quite analogous to the quantity $i$ occurring in equation (22) in the limiting case where the conductivity of the material of the wire becomes infinite.

*Methods for solving the fundamental equations*

In order to ascertain how the electromagnetic fields can propagate in vacuum when (perfect) conductors are present, two different courses can be followed.

One course is similar to that which we followed in the previous section: from equations (1) and (2), which have an integral form, by a suitable choice of integration contours we derive the Maxwell equations in the differential form, the equations (5) and (6) thus being a special case. The only point of difference, compared with what we did before, is that the electromagnetic field does not now depend upon one coordinate but upon all three. Once we have the Maxwell differential equations for the general case, then for any given configuration of conductors the problem is reduced to finding the solutions of a system of simultaneous differential equations with due observance of the boundary conditions (21) and (22). This method is perfectly general and in many cases in fact the only suitable one. It will be further explained, in connection with wave guides of a circular cross section, in Part II of this article (to be published in a subsequent number of this journal).

We shall first, however, follow another method which in principle is no less general though usually cumbersome. On the other hand, at least in simple cases, it expresses better what is physically essential in the solutions obtained. This method is based on the fact that the fundamental equations (1) and (2) are linear, *i.e.* they do not contain products of field components and consequently any linear combination of their solutions must also be a solution. For the unbounded, empty space this means that a situation where an arbitrary number of plane waves, with arbitrary direction and magnitude of the electric vector [7]), are propagated in arbitrary directions, corresponds to a solution of the fundamental equations. If there are (perfect) conductors present then this arbitrariness is considerably limited; only those superpositions of plane waves are permissible which satisfy the boundary conditions. For simple configurations of the conductors such permissible superpositions of plane waves can easily be found [8]). This will be illustrated in what follows with some examples.

**Propagation of the waves between parallel conducting planes and in rectangular wave guides**

Starting from the "superposition principle" outlined above we shall now consider how electromagnetic waves can be propagated with the three following configurations of conductors in vacuum;

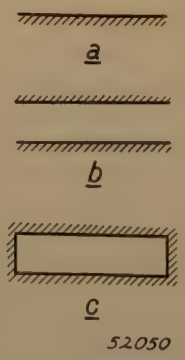



Fig. 5. Diagrammatic representation of the configuration of conductors dealt with in the text. *a*) one infinitely long and wide plane, *b*) two such planes parallel to each other, *c*) a rectangular wave guide, *i.e.* two pairs of planes as in (*b*), each pair being perpendicular to the other. In all three cases the plane of the drawing is perpendicular to the conducting planes.

[7]) By a "plane wave" is to be understood here and in the following pages a harmonic, linearly polarized, plane wave. For such a wave H is unambiguously determined by E and the direction of propagation.

[8]) The first to deal with the propagation of waves in wave guides as a problem of the superposition of plane waves was L. Brillouin (Rev. Gén. d'Electr. **40**, 227, 1936).

1) the space is bounded on one side, for instance underneath, by an infinitely long and wide, conducting plane (*fig. 5a*); 2) the space is bounded both underneath and on top, by two parallel planes (fig. 5*b*); 3) the space is bounded on four sides by two pairs of parallel planes, with the two pairs at right-angles to each other (fig. 5*c*). The last case represents a technically important example of a wave guide, namely that of a wave guide with a rectangular cross section.

*One single plane wave*

We begin by remarking that one single plane wave may be a permissible solution for the configurations of figs 5*a* and *b* but not for the wave guide of fig. 5*c*. In the cases of figs 5*a* and *b* it is indeed sufficient if the plane wave is chosen in such a way that its direction of propagation is parallel to the conducting planes and its electric vector is perpendicular to those planes. The boundary conditions (21) are then automatically satisfied, for then the normal component of **H** on the conducting planes and the component of **E** parallel to those planes are everywhere zero. From eq. (22) we then obtain immediately the values of the surface charge and surface current. Any other choice of the propagation direction of the plane wave or of the directions of **E** and **H** would not comply with the boundary conditions. Now it is readily seen that with the permitted choice of the plane wave one cannot add two other conducting planes (the case of fig. 5*c*) without coming into conflict with the boundary conditions (21) on these new planes. A single plane wave cannot, therefore, be propagated in a rectangular wave guide.

From this, however, one can draw farther-reaching conclusions. In the first place it is clear that a superposition of plane waves all having a direction of propagation parallel to the axis of the rectangular wave guide cannot produce an allowed wave either. In the second place the propagation of a purely transverse but possibly not plane wave in a rectangular wave guide — that is to say, a wave where both **E** and **H** have no component in the direction of the axis of the wave guide — is also impossible (because, as may be proved, this can be imagined as a superposition of plane waves all propagated parallel to the axis of the wave guide). The validity of this result is not confined to rectangular wave guides. So long as a wave guide consists of one conductor, that is to say so long as the boundary of its cross section consists of one continuous closed line (circle, ellipse, etc.), the fact remains that purely transverse waves cannot propagate in such wave guides. In the case of waves which can indeed be propagated in such wave guides the **E** and **H** fields always have a longitudinal component. On the other hand such does not hold hold for a wave guide consisting of two conductors, *i.e.* having a cross section bounded by two contours; in a coaxial cable for instance a purely transverse wave can indeed be propagated.

The situation described above, where a purely transverse plane wave is propagated between two infinite perfectly conducting planes, forms the basis for a rigorous treatment of a Lecher system consisting of two wide parallel metal strips at a short distance from each other.

It can easily be proved that with such a Lecher system the equations for the current and voltage are approximately equivalent to the differential equations for a plane wave in the previous section. It is only necessary to define the voltage in an unambiguous manner, since the current in the Lecher system is nothing else than the total surface current calculated from the boundary condition. We shall not go into that any further here, because a similar problem will be dealt with in Part II, *viz.* the rigorous deduction of the equations applying for a Lecher system consisting of two coaxial cylinders (the coaxial cable). We would only mention that the ratio $E^\circ/H^\circ = \sqrt{\mu/\varepsilon}$, for vacuum approximately equal to 377 ohms (see eq. (17)), then assumes the significance of the characteristic impedance of the Lecher system. For this reason one sometimes calls the quantity $\sqrt{\mu/\varepsilon}$ the characteristic impedance of the medium for plane waves.

*Superposition of two plane waves*

We shall now superpose one upon the other two plane waves having different directions of propagation and different **E** vectors but equal lengths $E$ of these vectors and equal frequency $\nu = \omega/2\pi$. We shall show that by doing this in a suitable manner we find electromagnetic waves (of course not plane and not transversal) capable of propagation in the rectangular wave guide.

Obviously any combination of plane waves that is not allowed for the case of fig. 5*a* (or 5*b*) will certainly not be allowed for the case of fig. 5*b* (or 5*c*). We shall therefore first consider the simple case of fig. 5*a*.

Let us consider a plane wave that strikes the conducting plane at an angle in such a way that **E** is parallel to the plane. We shall call this wave the "incident wave". Let $\vartheta$ (see *fig. 6a*) represent the angle between the direction of propagation of this wave and the conducting plane. From what has been said above it follows that this wave is not an allowed solution of the problem even for $\vartheta = 0$. But by superposing on the incident wave a second plane wave, the "reflected wave", we can easily

construct an allowed solution. It is evident that in order to comply with the boundary condition $E_t = 0$ on a conducting plane we must choose $\mathbf{E}^{\text{refl}} = -\mathbf{E}^{\text{inc}}$ ("refl" and "inc" refer here and in the

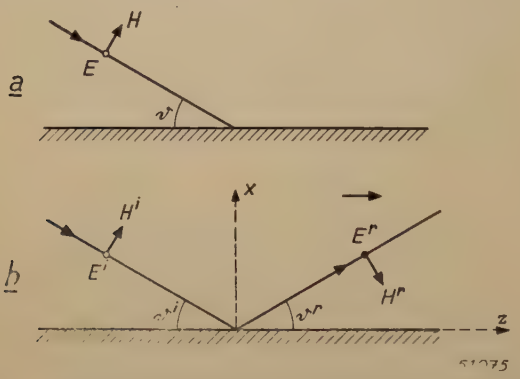



Fig. 6. Direction of propagation and position of the **E** and **H** vectors of the plane waves considered in the text, striking a conducting plane and reflected from it: *a*) the incident wave, *b*) the incident and the reflected wave (the arrow to the right of the illustration indicates the direction of the resulting wave). The **H** vectors lie in the plane of the drawing; the **E** vectors are at right-angles to it. ● signifies that an **E** vector is directed towards the reader, ○ that it has the opposite direction. The indices *i* and *r* refer to the incident and reflected waves respectively.

following to the reflected and incident waves respectively); both waves being plane, $\mathbf{E}^{\text{refl}}$ and $\mathbf{E}^{\text{inc}}$ will then everywhere be parallel. Having done this, we only have to choose $\vartheta^{\text{refl}} = \vartheta^{\text{inc}}$ in order to comply with the boundary condition $H_n = 0$. In this manner we arrive at the situation diagrammatically represented in fig. 6*b*. From simple considerations of symmetry it is not difficult to understand that the wave resulting from the superposition of the incident and reflected waves is propagated in a direction parallel to the conducting plane and perpendicular to $\mathbf{E}^{\text{inc}}$ (or $\mathbf{E}^{\text{refl}}$); in fig. 6*b* this direction is indicated by an arrow. Now what is the form of this wave? In order to answer this question fully the simplest way is to perform analytically the superposition of the incident and reflected waves, i.e. to write the expressions

$$\mathbf{E} = \mathbf{E}^{\text{inc}} + \mathbf{E}^{\text{refl}}, \qquad \mathbf{H} = \mathbf{H}^{\text{inc}} + \mathbf{H}^{\text{refl}}.$$

If the system of coordinates $x, y, z$ is chosen as indicated in fig. 6*b* (the equation of the conducting plane is then $x = 0$), then a simple calculation gives the following expressions

$$\left.\begin{aligned} E_y &= A \sin(xk \sin\vartheta)\, e^{j(\omega t - zk\cos\vartheta)}, \\ H_x &= A \sqrt{\frac{\varepsilon_0}{\mu_0}} \cos\vartheta \sin(xk\sin\vartheta)\, e^{j(\omega t - zk\cos\vartheta)}, \\ H_z &= A \sqrt{\frac{\varepsilon_0}{\mu_0}} \sin\vartheta \cos(xk\sin\vartheta)\, e^{j(\omega t - zk\cos\vartheta - \pi/2)}, \\ E_x &= E_z = H_y = 0 \end{aligned}\right\} \quad (23)$$

($A$ = an arbitrary constant, $k = \omega\sqrt{\mu_0\varepsilon_0} = \omega/c = 2\pi/\lambda$, where $\lambda$ is the wavelength of each of the plane waves.)

This shows that the resultant wave is indeed propagated parallel to the conducting plane and perpendicular to its **E** vector. Its phase velocity $v$ is

$$v = \omega/k \cos\vartheta = c/\cos\vartheta,$$

thus greater than $c$, whilst its wavelength $\lambda_z$ is

$$\lambda_z = v/\nu = c/\nu \cos\vartheta = \lambda/\cos\vartheta\,. \qquad (24)$$

The form of the wave (23) is rather complicated. **E** and **H** are at right-angles to each other, as is the case with a plane wave, but whereas the **E** field is transverse the **H** field is not so: contrary to the case of a plane wave, here the magnetic field has a component in the direction of propagation. Such a wave is called a "transverse electric wave" (abbreviated "TE" wave) or "H-wave". Furthermore we see that the wave is no longer linearly polarized: the direction of **H** at each point of the space changes with the time, owing to $H_x$ and $H_z$ differing 90° in phase. Finally, we find that the wave is no longer plane either. The values of the E, H fields in a plane perpendicular to the direction of propagation ($z$-direction) depend upon $x$: for $z$ = constant $H_x$, $H_z$, $E_y$ vary sinusoidally with $x$, that is to say, along the $x$-direction the wave (23) behaves as a standing wave. In particular $H_x = 0$ and $E_y = 0$ (regardless of the value of $t$) in each plane which is parallel to the conducting plane and for which

$$x = \frac{m\pi}{k \sin\vartheta} = \frac{m\lambda}{2 \sin\vartheta} \quad (m = 0, 1, 2\ldots) \quad (25)$$

applies. If we introduce the wavelength

$$\lambda_x = \lambda/\sin\vartheta \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (26)$$

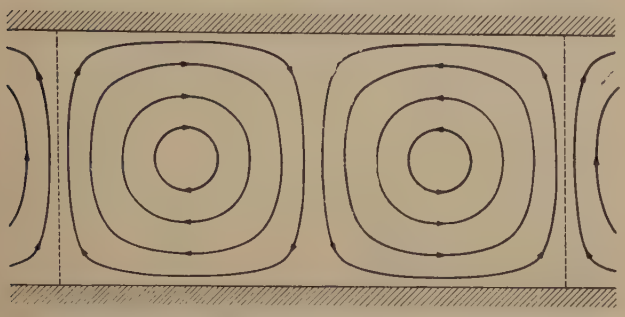
of the said standing wave, we can write instead of (25):

$$2x = m\lambda_x \quad (m = 0, 1, 2\ldots), \quad . \quad . \quad . \quad (27)$$

which is self-explanatory.

The consequence of this is highly important. When a second conducting plane is applied above and parallel to the conducting plane $x = 0$ at a distance $x$ given by (27) then the boundary conditions $E_t = H_n = 0$ are automatically satisfied also for this second plane; this means that the wave (23) is also an allowed solution of the problem of fig. 5*b* provided the distance between the conducting planes complies with (27). We can, however, verify the correctness of a still farther-reaching statement that the wave (23) — still under the condition (27) — is also an allowed solution of the problem of the rectangular

wave guide. The fact is that if we apply two other conducting planes at an arbitrary distance from each other parallel to the direction of propagation of the wave (23) and perpendicular to its



52094

Fig. 7. The $H$-lines in the case of an $H_{10}$-wave in a rectangular wave guide ("instantaneous view"); a cross section parallel to one pair of walls (thus also parallel to the direction of propagation). The $E$-lines are perpendicular to the plane of the drawing. The $H$-lines are given by the equation

$$\sin 2\pi\xi \cdot \sin 2\pi\zeta = \text{constant},$$

where $\xi = x/\lambda_x$ and $\zeta = z/\lambda_z$. The distance between the two dotted vertical lines is equal to $\lambda_x$; the distance between the two horizontal lines representing the walls is equal to $\lambda_x/2$. The $H$-lines in the case of an $H_{m0}$ wave are obtained by drawing the picture for the $H_{10}$-wave $m$ times one above the other.

E vector (having the direction of the $y$-axis), then the boundary conditions (21) are likewise automatically satisfied at these two new conducting planes. Different values of $m$ thus correspond to different forms of the wave ($m + 1$ is equal to the number of "nodal planes" of the standing wave along the $x$-direction). The H-waves of the type (23) are therefore denoted by $H_{m0}$; the second index 0 reminds us that the wave is not dependent on the $y$-coordinate.

The magnetic lines of force (for $t$ = constant) of the $H_{10}$-wave are represented in *fig. 7*; $E$-lines are perpendicular to the plane of the drawing. In *fig. 8* we have a diagram representing the surface

52095

Fig. 8. Representation of the density distribution of the surface current on the walls of a rectangular wave guide through which an $H_{10}$-wave is propagated ("instantaneous view"). The width of the "lines of flow" perpendicular to the magnetic lines of force in fig. 6 is proportional to the current density and also — owing to the boundary condition (22) — proportional to the value of the magnetic field strength at the wall.

current density which follows from the boundary conditions (22).

For the above-described "construction" of the $H_{m0}$-waves we have taken a plane wave of which the electric vector **E** is parallel to the conducting plane and which strikes the conducting plane at an angle. If such a construction had been built up from an oblique incident plane wave whose magnetic vector **H** was parallel to the conducting plane, then we should have obtained in precisely the same way a wave — the "transverse magnetic" (TM-wave) or "E-wave", *viz.* an "$E_{m0}$-wave" — which could be propagated between two parallel planes (fig. 5*b*) but not in a rectangular wave guide (fig. 5*c*). From the analogy between an $E_{m0}$-wave and an $H_{m0}$-wave it immediately follows that, just as for an $H_{m0}$-wave **E** has the direction of the $y$-axis and does not depend upon $y$ (see eq. (23)), for an $E_{m0}$-wave **H** must also have the direction of the $y$-axis and cannot be dependent upon $y$. Therefore, the two walls of the wave guide which are perpendicular to the $y$-axis could not possibly be applied in such a way as to satisfy the boundary condition $H_n = 0$ everywhere along those walls. Consequently it is not possible to obtain permissible E-waves in rectangular wave guides by the superposition of two plane waves. This would require the superposition of at least four plane waves. The E-waves are then found to be of a more complicated form: they behave as standing sinusoidal waves not only along the $x$-direction but also along the $y$-direction, and that is why they are called "$E_{mn}$-waves". By a suitable superposition of four plane waves it is possible to construct also the H-waves analogous to "$E_{mn}$-waves", called the "$H_{mn}$-waves".

*The cut-off frequency*

In the foregoing we have adapted, as it were, the position of the conducting planes to the form of the waves (23). From this one might be inclined to conclude that an $H_{m0}$-wave can only be propagated in a rectangular wave guide when there is a discontinuous series of values for the distance between two parallel walls. However, such a conclusion is incorrect. If the said distance is $a$ then in eq. (25) we must put $x = a$ and determine from that the angle $\vartheta$, which, just like $m$, plays the part of a parameter:

$$\sin\vartheta = \frac{m\lambda}{2a} \quad (m = 0, 1, 2 \ldots .) \quad . \quad . \quad . \quad (28)$$

With the aid of eq. (28) we can eliminate $\vartheta$ from eq. (23) and then obtain the following expressions for the $H_{m0}$-wave:

$$\left.\begin{aligned} E_y &= A \sin\frac{m\pi x}{a}\, e^{2\pi j\varphi}, \\ H_x &= A \sqrt{\frac{\varepsilon_0}{\mu_0}} \sqrt{1 - \left(\frac{m\lambda}{2a}\right)^2} \sin\frac{m\pi x}{a}\, e^{2\pi j\varphi}, \\ H_z &= A \sqrt{\frac{\varepsilon_0}{\mu_0}} \left(\frac{m\lambda}{2a}\right)^2 \cos\frac{m\pi x}{a}\, e^{2\pi j(\varphi - 1/4)}, \\ E_x &= E_z = H_y = 0, \end{aligned}\right\} \quad (29)$$

where

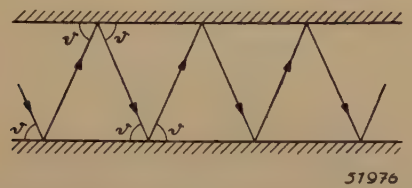
$$\varphi = \nu\left\{t - \frac{z}{c}\sqrt{1-\left(\frac{m\lambda}{2a}\right)^2}\right\} \quad . \quad . \quad . \quad (30)$$

However, it is not always possible to determine a real angle $\vartheta$ from eq. (28), because a real angle $\vartheta$ can only be found from this equation when its right-hand side is $\leqq 1$, that is to say only if $\lambda$ is not greater than a certain wavelength which is called the cut-off wavelength $\lambda_c$ and which is given by

$$\lambda_c = \frac{2a}{m} \quad (m = 1, 2 \ldots.); \quad . \quad . \quad . \quad (31)$$

The frequency $\nu$ must therefore be greater than the cut-off frequency $\nu_c = mc/2a$.

If $\lambda$ satisfies the condition $\lambda > \lambda_c$, that is to say if $\vartheta < 90°$, then the $H_{m0}$-wave may also be regarded as one plane wave (wavelength $\lambda$) reflected to and fro between a pair of parallel conducting planes and propagated in this zig-zag fashion in the $z$-direction



51976

Fig. 9. The zig-zag course of a plane wave between two planes the situation where the angle $\vartheta$ is almost 90° corresponds to a wave whose length is only a little smaller than the cut-off wavelength.

(see *fig. 9*). This interpretation leads to the conclusion that when $\lambda = \lambda_c$, *i.e.* $\vartheta = 90°$, the wave can no longer be propagated in the $z$-direction. We shall now verify this with the aid of the formulae given above, when it will at the same time be made clear what happens when $\lambda > \lambda_c$. Since, owing to (28) and (31),

$$\cos\vartheta = \sqrt{1-(m\lambda/2a)^2} = \sqrt{1-(\lambda/\lambda_c)^2}, \quad . \quad (32)$$

the $z$-dependence of the wave (29)-(30) is given by

$$e^{-2\pi j z\sqrt{(1/\lambda)^2-(1/\lambda_c)^2}}. \quad . \quad . \quad . \quad . \quad . \quad (33)$$

For $\lambda = \lambda_c$ the exponent becomes zero, that is to say the wave (29) is in fact no longer dependent upon $z$; it is only a standing wave along the $x$-direction. But the expression (33) shows us something more, *viz.* that for $\lambda > \lambda_c$ the wave (29) is damped because then the expression underneath the radical is negative, that is to say the exponent is a real number. For $\lambda > \lambda_c$ (29) thus represents a standing wave along the $x$-direction whose amplitude depends exponentially upon $z$.

The practical consequences of the existence of the cut-off frequency are now clear. An $H_{m0}$-wave cannot be propagated in a rectangular wave guide if its frequency $\nu$ is less than $\nu_c$: if such a wave is admitted at the entrance of the wave guide it changes into a standing wave (perpendicular to the axis of the wave guide) whose amplitude decreases exponentially along the axis of the wave guide, so that at some distance from the entrance it can hardly be perceived.

This result can also be interpreted in the following manner. From (27) and (31) it is directly deduced that

$$\lambda_c = \lambda_x \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (34)$$

One may therefore say that a wave with $\lambda > \lambda_c$ cannot be propagated in the wave guide "because" its wavelength is too great to be able to form a standing wave in the $x$-direction which matches the given values of $a$ and $m$.

From the foregoing it is evident that the damping of the standing wave (for $\lambda > \lambda_c$) is not an absorption phenomenon, for if the walls of the wave guide are perfectly conducting no heat is developed.

When we have to do with $H_{mn}$ or $E_{mn}$ waves, which therefore represent standing waves both in the $x$-direction and in the $y$-direction, the cut-off wavelength is given by:

$$\frac{1}{\lambda_c} = \sqrt{\left(\frac{1}{\lambda_x}\right)^2 + \left(\frac{1}{\lambda_y}\right)^2},$$

where

$$\lambda_x = \frac{2a}{m}, \qquad \lambda_y = \frac{2b}{n};$$

$a$ and $b$ being the lengths of the sides of the rectangular cross section of the wave guide. From this we see that the $H_{10}$-wave ($m = 1$, $n = 0$) has the smallest cut-off frequency, *viz.* (assuming $a > b$):

$$\frac{c}{\nu_c} = \lambda_c = 2a. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (35)$$

Consequently the $H_{10}$-wave is very important from the practical point of view.

It can generally be proved that with any wave guide there is a cut-off frequency $\nu_c > 0$ for every wave that is not purely transverse. Now we have already seen that a purely transverse wave cannot be propagated in wave guides having a cross section bounded by one single continuous line (*e.g.* rectangular, circular or elliptical wave guides). Such wave guides act as filters, "cutting off" the low-frequency components of any electromagnetic wave entering them.

In this connection it may not be superfluous to stress the fact that in this article we are concerned only with the problem as to what modes of propagation of electromagnetic waves are possible in wave guides. Here we leave out of consideration the no less important but more difficult problem as to which of the possible waves will actually be propagated in a wave guide when, for instance, a radiating aerial is placed at the entrance of the wave guide.

### *Damping of the waves owing to losses*

In the foregoing we have assumed throughout that the conductors were perfect and that the medium through which the waves are propagated was vacuum. In actual wave guides the medium is mostly air and the walls are not perfectly conducting. Consequently waves are always attenuated in wave guides, for the air can lead to "dielectric losses" and the finite conductivity of the metal walls to the development of Joule heat. Under normal conditions the former effect is of course negligibly small compared with the latter.

Dielectric losses can be exactly calculated by superposing plane damped waves one upon the other in the manner described. As explained in a previous chapter, $k$ then becomes a complex number: $k = k_1 - jk_2$. The amplitudes of the field vectors of the resultant wave are each multiplied by a factor $e^{-k_2 z}$; $k_1$ plays the part which $k = \omega/c$ plays in vacuum. If the conductivity of the dielectric is very small then $k_1 \approx k$, so that the values of the cut-off frequency are hardly affected.

The damping of the waves caused by losses in the walls is calculated only approximately. It is assumed that this damping finds expression in the form of a factor $e^{-\alpha z}$ by which all field amplitudes are multiplied. We shall not go into the question how $\alpha$ is calculated but would only state that it is essential to know how the surface current is distributed over the walls. The losses are calculated by imagining this current to be "spread out" in a layer the thickness of which is equal to the "penetration depth" $d$.

# THE PART PLAYED BY OXYGEN AND NITROGEN IN ARC-WELDING

by J. D. FAST.

621.791.753:669.1.014.6:541.123.28

The action of oxygen and nitrogen on iron and steel is dealt with here at some length as an introduction to an article that will follow in which the function of the coating of welding electrodes will be analysed. This article is devoted to an estimation of the solubility of oxygen and nitrogen in liquid as well as in solid iron and to a discussion of the harmful and the useful effects of these elements in electric welding. As harmful effects are discussed the causation of porosity and the adverse influence upon mechanical strength, whilst the useful effects discussed are the promotion of transfer of heat and transport of the metal across the arc. This leads to a discussion also of the part played by the reaction between oxygen and carbon.

When welding by the commonly applied method of Slavianoff an arc is struck between the workpiece to be welded and a metal rod called the electrode. The latter gradually melts during the welding and thus functions not only as electrode but also as filler rod.

At first no steps were taken to protect the molten metal against the attack of the oxygen and the nitrogen in the air, with the result that very poor quality welds were obtained. When unprotected rods of mild steel containing for instance 0.1% C, 0.1% Si and 0.4% Mn are used it is found that in the transfer of the metal to the workpiece the content of these elements is greatly reduced by oxidation. On the other hand the originally very small content of oxygen and nitrogen is increased to something like 0.15% N and 0.25% O, as a chemical analysis of the bead shows. Accordingly a microscopic examination of the deposited metal may show the presence of a considerable quantity of oxide and nitride. Furthermore, there appear to be a number of cavities in the metal due to the formation of carbon monoxide through oxidation of the carbon contained in the steel and, as will be demonstrated, also to the release of nitrogen during the cooling and solidifying of the metal.

In order to obtain welds with satisfactory mechanical properties it is therefore essential to protect the metal against the attack of oxygen and nitrogen while being transferred from the electrode to the workpiece.

In principle the simplest way to provide such a protection is to carry out the welding in a gas that is absolutely free of components which react with iron. Welding in hydrogen for instance has become well-known; use is thereby made of the large quantity of heat released by the recombination of the atomic hydrogen formed in the arc. This method, however, is a rather expensive one and is consequently used only in special cases. The same applies to welding in helium and argon, as recently developed.

A more economical protection is obtained by coating the electrodes with substances which during the welding process keep the oxygen and nitrogen away from the metal, either by the development of large quantities of other gases (organic substances) or by the formation of a sealing slag on the metal (mineral substances). Frequently a combination of both these processes is applied.

In our next article we shall go fully into the protection afforded by a slag, but first of all it is necessary to study more closely the effect of oxygen and nitrogen on iron and steel, as will be done in this article. The obsolete method of welding with bare electrodes will only be referred to in so far as it will help to give a better understanding of the part played by the coating.

We shall concentrate our attention particularly upon the questions that are of importance in the technique of welding, viz:

a) The solubility of oxygen and nitrogen in liquid iron between its melting point and boiling point; this gives us an idea of the maximum quantities that may be absorbed while welding.

b) The solubility of oxygen and nitrogen in solid iron, particularly alpha iron (the form of iron which is stable below 910 °C and is body-centred cubic); this helps to gain an insight into the effect of these elements upon the mechanical properties of welds.

c) The chemical reaction between oxygen and carbon dissolved in liquid iron; this reaction is apt to cause porosity in the welds and, what is still more important, provides the propelling force for the transfer of the metal from the electrode to the workpiece.

The injurious as well as the useful effects of oxygen and nitrogen in the welding process will be discussed in the latter part of this article.

## The iron-oxygen system

### *Solubility of oxygen in liquid and in solid iron*

If liquid iron is attacked by oxygen and the pressure of the oxygen exceeds a certain value then all the metal will be gradually transformed into liquid oxide. This can be read from the partial constitutional diagram of the iron-oxygen system (*fig. 1*), when it is borne in mind that to each percentage of oxygen and each temperature there corresponds a certain partial pressure of the oxygen.

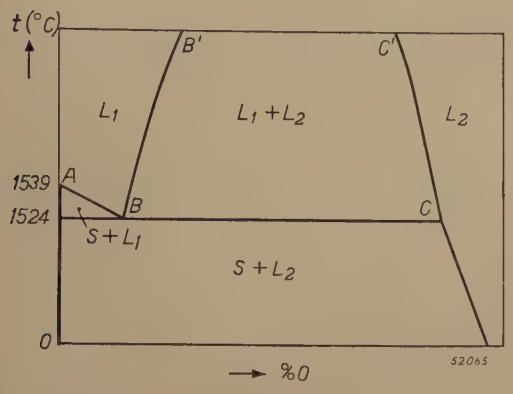



Fig. 1. Part of the constitutional diagram of the iron-oxygen system represented diagrammatically. $S$ signifies solid iron, $L_1$ liquid iron with oxygen in solution, $L_2$ liquid iron oxide of variable composition. The line $BB'$ is the solubility curve of oxygen in liquid iron. The solubility of oxygen in solid iron is so small that the solubility curve practically coincides with the $t$-axis. The position of the points $C$ and $C'$ does not differ much from the stoichiometric composition FeO.

Let us suppose that while maintaining a constant temperature (say 1600 °C) we cause the oxygen pressure above the liquid iron to rise gradually from zero. The state of saturation, indicated in the diagram by the solubility curve $BB'$, will then be reached already at a very low oxygen pressure. At 1600 °C this pressure is about $10^{-8}$ atm. As soon as the state of saturation has been reached, then, in addition to the liquid metallic phase, a liquid oxide phase begins to form, the composition of which is given by the position of the line $CC'$. According to Gibbs' phase rule the pressure should remain constant (about $10^{-8}$ atm.) until all the metal has been transformed into this oxide phase.

If the number of independent components in a system is $C$ and the number of phases $P$ then according to Gibbs' phase rule the number of variables (degrees of freedom) $F$ required to determine fully the state of the system is given by $F=C-P+2$. In the case in question for the area bounded on the left and right by the lines $BB'$ and $CC'$ we therefore have $F = 2-3+2 = 1$. Consequently in this area the state of the system is a function of only one variable, *e.g.* the pressure or the temperature. If we take the temperature for this variable then the pressure is a function only of the temperature. Consequently, when the value of the temperature is fixed, as we have done in this case by supposing the temperature to be constant at 1600° C, then also the pressure is fixed.

If at the constant temperature of 1600 °C the pressure of the oxygen is caused to rise further, then the liquid oxide, the composition of which on the line $CC'$ was not far removed from the stoichiometric composition FeO[1]), gradually absorbs more oxygen. Finally at a pressure of one atmosphere an oxygen content is reached which is already greater than that corresponding to the formula $Fe_3O_4$.

In electric arc welding with bare electrodes the reaction time is so short that actually not all the metal is transformed into oxide, as is required by the thermodynamic equilibrium, but only a partial oxidation takes place. Immediately underneath the electrode a molten mass is formed, called the pool, consisting of a metallic phase containing oxygen and covered by a thin layer of liquid oxide. At the interface the compositions of these two phases will be as represented by the lines $BB'$ and $CC'$ in fig. 1. The average composition of the metallic phase, however, will be given by points a little to the left of $BB'$ and that of the oxide phase by points a little to the right of $CC'$.

What is of particular importance in welding technique is the position of the line $BB'$, i.e. the solubility of oxygen (or FeO) in liquid iron as a function of the temperature [2]). For the temperature range between 1800 °K and 2083 °K this solubility has been accurately determined by Chipman and Fetters [3]).

It is highly important to know the solubility also at higher temperatures, for there are indications that while passing from the electrode to the workpiece the metal is in many cases heated to the boiling point of iron. According to most experimental data the boiling point of iron lies somewhere between 2700 and 2800 °K.

By means of a few simple hypotheses regarding the thermal effect and the change in entropy taking place when liquid Fe is homogeneonsly mixed with liquid FeO, the following relation between solubility and temperature can be deduced [4]):

$$T = \frac{C}{R} \frac{1-2x}{\ln \frac{1-x}{x}} \quad \ldots\ldots \quad (1)$$

---

1) According to some investigators the line $CC'$ lies to the left of the composition FeO whilst others place it to the right.

2) Of course the solubility can be expressed in % FeO as well as in % O.

3) J. Chipman and K. L. Fetters, Trans. Am. Soc. Metals **29**, 953-967, 1941.

4) Cf. J. D. Fast, Philips Res. Rep. **2**, 205-227, 1947.

$T$ being the absolute temperature and $x$ the solubility of FeO expressed as a molecular fraction, i.e. the number of molecules of FeO in the saturated solution divided by the total number of molecules. $R$ is the gas constant and $C$ a constant that has to be determined experimentally.

Formula (1) indicates that when plotting $(1-2x)/\log[(1-x)/x]$ as a function of $T$ we should obtain a straight line through the origin. As shown in *fig. 2*, Chipman's and Fetters' measurements

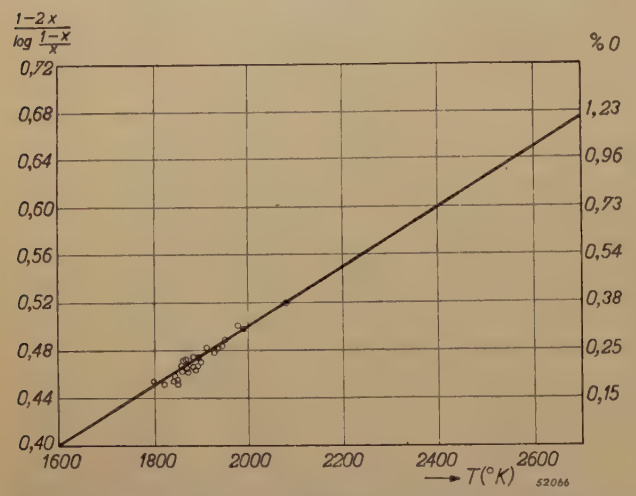



Fig. 2. The expression $(1-2x)/\log[(1-x)/x]$, where $x$ indicates the solubility (expressed as molecular fraction) of FeO in liquid iron, as a function of the absolute temperature. To the right of the figure the values of solubility are given in wt%. The theoretical straight line (equation (1)) points to the invisible origin at 0°K. The experimental points (small circles) have been taken from the measurements of Chipman and Fetters (see footnote [3])).

satisfy this requirement quite well. The straight line in the diagram has been drawn in such a way as to point precisely towards the invisible origin at O°K. From its slope we find for the constant $C$ in (1) the value

$$C = 18\,300 \text{ cal.}$$

The satisfactory agreement between theory and experiment encourages us to use equation (1) also for calculating the solubility at higher temperatures. These (calculated) solubilities can be read directly from fig. 2.

The solubility of oxygen in solid iron is so small that it could not be determined experimentally. In fig. 1 the solubility curve coincides with the $t$-axis.

### *Absorption of oxygen when welding with bare electrodes*

Since the boiling point of iron lies at about 2750 °K, from fig. 2 it may be expected that when welding with bare electrodes the molten metal will always contain less than 1.2% of oxygen.

Of interest in this connection are the experiments carried out by Losana [5]). He used bare electrodes of different diameters and made of different kinds of steel. It was found that both the oxygen and the nitrogen content of the deposited metal increased as the thickness of the electrode decreased. In not a single experiment was a content found lower than 0.140% O, or higher than 0.953% O. The results of these experiments, therefore, are not in contradiction with fig. 2.

As an example taken from a series of experiments by Losana we give in *table I* the C, Mn, Si, P, S, O and N contents of the metal deposited from electrodes containing 0.10% C, 0.89% Mn, 0.17% Si, 0.015% P and 0.021% S (the balance being Fe).

**Table I.** Composition of the metal deposited in some of Losana's experiments [5]) when welding with bare electrodes containing 0.10% C, 0.89% Mn, 0.17% Si, 0.015% P and 0.021% S.

| Wire dia. mm | C % | Mn % | Si % | P % | S % | O % | N % |
|---|---|---|---|---|---|---|---|
| 1 | ? | ? | ? | ? | ? | 0.720 | 0.218 |
| 2 | 0.026 | 0.054 | 0.039 | 0.010 | 0.019 | 0.550 | 0.180 |
| 4 | 0.048 | 0.070 | 0.053 | 0.014 | 0.021 | 0.302 | 0.130 |
| 6 | 0.068 | 0.124 | 0.058 | 0.014 | 0.022 | 0.140 | 0.105 |

For a more exact comparison with the theory the experiments should have been carried out with wire of pure iron. In the experiments referred to C, Mn and Si were, it is true, oxidized to a condiderable extent but by no means completely.

The drop in the oxygen content with increasing diameter of the electrode wire is probably related to the fact that the thicker the wire the larger are the droplets of metal transferred from the electrode to the workpiece. It may be that consequently the temperature of the metal crossing the arc is lower when thicker rods are used. Moreover, we have to take into account the fact that there is no time for the equilibrium of dissolution and the chemical equilibria to adjust themselves fully. In the following section we shall see from the example of the reaction between oxygen and carbon that these equilibria are in fact unable to adjust themselves in the process of welding.

### The reaction between oxygen and carbon in electric arc welding

To a liquid iron phase containing small quantities of C and FeO (or O) in solution, in the state of

[5]) L. Losana, Metallurgia Italiana **26**, 391-403, 1934.

equilibrium there belongs a gas phase consisting of a mixture of CO and $CO_2$. The reactions that can take place between the components in the two phases are the following:

$$[FeO] + [C] \rightleftarrows [Fe] + CO \quad . . . . . \quad (2)$$

$$2\,[FeO] + [C] \rightleftarrows 2[Fe] + CO_2 \quad . . . . . \quad (3)$$

$$[FeO] + CO \rightleftarrows [Fe] + CO_2 \quad . . . . . \quad (4)$$

$$[C] + CO_2 \rightleftarrows 2\,CO \quad . . . . . \quad (5)$$

where the brackets indicate the components of the homogeneous liquid phase. Two of these equations, however, are sufficient to describe the equilibrium between the liquid and the gaseous phases, since (4) is the difference between (3) and (2), and (5) is equal to twice (2) less (3).

With the aid of thermodynamics and experimental data taken from literature it is possible to calculate the positions of the said equilibria as functions of the temperature of the steel (see the article quoted in footnote [4]). Without going into these calculations we give here in *table II* the equilibrium pressures of CO and $CO_2$ which should correspond to the C and O contents of the deposited metal given in table I. The calculations have been carried out for two temperatures, *viz:* 2300 °K and 2700 °K.

**Table II.** Calculated equilibrium pressures of CO and $CO_2$ (in atmospheres) corresponding to the C and O contents in table I.

| % C | % O | 2300 °K | | 2700 °K | |
|---|---|---|---|---|---|
| | | $P_{CO}$ | $P_{CO_2}$ | $P_{CO}$ | $P_{CO_2}$ |
| 0.026 | 0.550 | 10.4 | 1.23 | 14.3 | 0.75 |
| 0.048 | 0.302 | 11.3 | 0.79 | 15.3 | 0.47 |
| 0.068 | 0.140 | 7.7 | 0.26 | 10.3 | 0.15 |

The choice of 2300°K as lowest temperature is due to the fact that the highest oxygen content of deposited metal in table I of which also the C content is known is 0.55%, which according to fig. 2 would correspond already to a temperature of about 2250°K of the metal being transferred if the equilibrium between liquid iron and liquid ferro-oxide had indeed been established during the deposition process. The actual temperature was probably higher.

In the calculations which led to the pressures given in table II extrapolations were employed which may have caused errors of 10% in the results. This, however, does not detract from the conclusion obviously to be drawn from table II that when welding with bare electrodes conditions are far removed from the state of chemical equilibrium. The $(CO + CO_2)$-pressure cannot in actual fact have amounted to more than a fraction of one atmosphere.

This failure to reach states of equilibrium will prove to be of essential importance also when dealing with coated electrodes.

**The iron-nitrogen system**

*Dissociation pressure of iron nitride*

When iron was attacked by oxygen at high temperatures saturated solutions were obtained, as shown above, at $O_2$-pressures far below the partial pressure of this gas in the air. In the case of nitrogen the reverse is found: at high temperatures the state of saturation both for liquid and for solid iron is reached only at $N_2$-pressures far above the partial pressure of nitrogen in air. Consequently it has not yet been possible to reach the saturation concentrations for liquid iron. As we shall see farther on, it has however been possible to determine the concentrations corresponding to a nitrogen pressure of 1 atm.

For alpha iron [6]) it was possible to reach the condition of saturation by an artifice. Instead of $N_2$, mixtures of $NH_3$ and $H_2$ were caused to react with iron. As soon as the limit of solubility is exceeded a new phase is formed (the nitride $Fe_4N$), the dissociation pressures of which at various temperatures naturally correspond to the equilibrium pressures of the saturated solutions of nitrogen in iron. As we shall see below, these dissociation pressures can be calculated and will enable us to compute an upper limit for the solubility of nitrogen in alpha iron. We shall also see (table III) that more direct determinations of the position of the solubility line *AC* diagrammatically represented in *fig. 3* give greatly divergent values.

The explanation why $NH_3$ is so much more active than $N_2$ lies in the fact that the equilibrium

$$8\,Fe + 2\,NH_3 \rightleftarrows 2\,Fe_4N + 3\,H_2 \quad . . . \quad (6)$$

is reached comparatively quickly, whereas the reactions

$$N_2 + 3\,H_2 \rightleftarrows 2\,NH_3 \quad . . . . . \quad (7)$$

and

$$2\,Fe_4N \rightleftarrows 8\,Fe + N_2 \quad . . . . . \quad (8)$$

take place but very slowly. If, for instance, $NH_3$ is heated for a long time to 500 °C it dissociates almost entirely into hydrogen and nitrogen, since the

[6]) Alpha iron is the body-centred cubic form of iron which is stable below 910 °C. The modification of iron which is stable between 910 and 1400 °C and is face-centred cubic is called gamma iron. The body-centred cubic form is again stable between 1400 °C and the melting temperature (1540 °C) and this is now called delta iron.

stoichiometric gas mixture in the state of equilibrium at 500 °C and 1 atm contains only 0.12 vol % $NH_3$. When, however, iron is heated to 500 °C in a stream of $NH_3 + H_2$ the velocity of the gas can

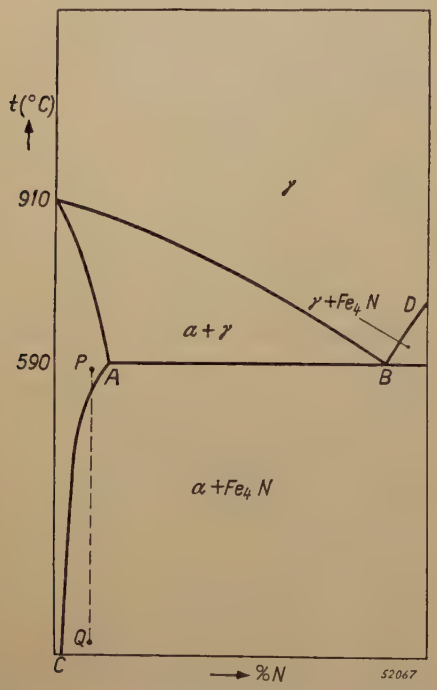



Fig. 3. Part of the constitutional diagram of the iron-nitrogen system represented diagrammatically. Here $\alpha$ and $\gamma$ indicate respectively alpha iron and gamma iron. The solubility curve of nitrogen in solid iron is given by the lines *CA* and *BD*.

easily be chosen such that the percentage of $NH_3$ drops only very little. The iron then reacts with a mixture of $NH_3$ and $H_2$, which corresponds to very high nitrogen pressures in the state of equilibrium. By way of illustration, the constant of the reaction (7)

$$K_{NH_3} = \frac{p^2_{NH_3}}{p_{N_2} \cdot p^3_{H_2}} \quad \ldots\ldots \quad (9)$$

at 500 °C amounts to $1.5 \cdot 10^{-5}$. A gas mixture comprising 70 vol % $NH_3$ + 30 vol % $H_2$ of 1 atm thus corresponds at 500 °C to a nitrogen pressure

$$p_{N_2} = \frac{10^5}{1.5} \cdot \frac{(0.7)^2}{(0.3)^3} = 1.2 \cdot 10^6 \text{ atm},$$

whilst a mixture of 98 % $NH_3$ + 2 % $H_2$ corresponds even to a nitrogen pressure of about $8 \cdot 10^9$ atm [7]).

When mixtures of $NH_3$ and $H_2$ of varying composition are passed over iron at 500 °C then the formation of $Fe_4N$ begins as soon as the $NH_3$ content rises above 20 vol %. The reaction takes place in the opposite direction as soon as the $NH_3$ content falls below 20 %. If we now substitute $p_{H_2} = 0.8$ and $p_{NH_3} = 0.2$ in (9) then we find that $p_{N_2} \approx 5200$ atm. Considering (7) and (8) this means that also the dissociation pressure of $Fe_4N$ in equilibrium with Fe amounts to about 5200 atm $N_2$ at 500 °C.

In the same way we can determine the dissociation pressure for other temperatures, and from the experimental data available we deduce the following relation between the dissociation pressure of $Fe_4N$ (in equilibrium with Fe) and the absolute temperature:

$$\log p = -\frac{1760}{T} + 5.99 \quad \ldots\ldots \quad (10)$$

According to this formula the dissociation pressure at 20 °C is already about 1 atm. The $Fe_4N$ needles in steel, so well known in metallography, should, therefore, dissociate spontaneously; the fact that they continue to exist is due only to the inertia of this dissociation. It is the same as with iron carbide (cementite) $Fe_3C$, which in iron-carbon alloys should really dissociate spontaneously into iron and graphite but only does so at an imperceptibly low rate.

*Solubility of nitrogen in solid iron*

Some research workers understand by the solubility of nitrogen in iron something different from what we have understood it to be in the foregoing, in conformity with the conventional definition [8]). They understand it to be the amount of nitrogen contained in the metal in equilibrium with nitrogen of 1 atm. We have seen, however, that the true solubility say at 500 °C corresponds to a pressure of about 5200 atm. Now in order to be able to speak about the solubility at a certain nitrogen pressure without causing confusion we shall from now onwards call the (true) solubility as indicated in the constitutional diagram of fig. 3 by the lines *CA* and *BD* the maximum solubility.

Reliable data in respect to the solubility of nitrogen in solid iron are only available for a pressure of 1 atm. They are given by Sieverts and his collaborators [9]) and are represented by *fig. 4*.

According to Sieverts the solubility in alpha iron is 0.002 wt % at 890 °C and 0.0004 wt % at 750 °C. At these relatively low temperatures equilibrium is established so slowly that the values given

[7]) Of course these figures are not exact, because equation (9) only strictly applies as long as the gases behave as ideal gases. It would be better to substitute "nitrogen activity" for "nitrogen pressure".

[8]) This reads: solubility is the maximum quantity that can be absorbed without a new phase beginning to form.

[9]) A. Sieverts G. Zapf and H. Moritz, Z. physik. Chem. A **183**, 19-37, 1938.

are less reliable than those for gamma and delta iron (see note 6). Now the iron atoms in alpha as well as in delta iron form a body-centred cubic lattice and it is therefore to be expected that one continuous curve

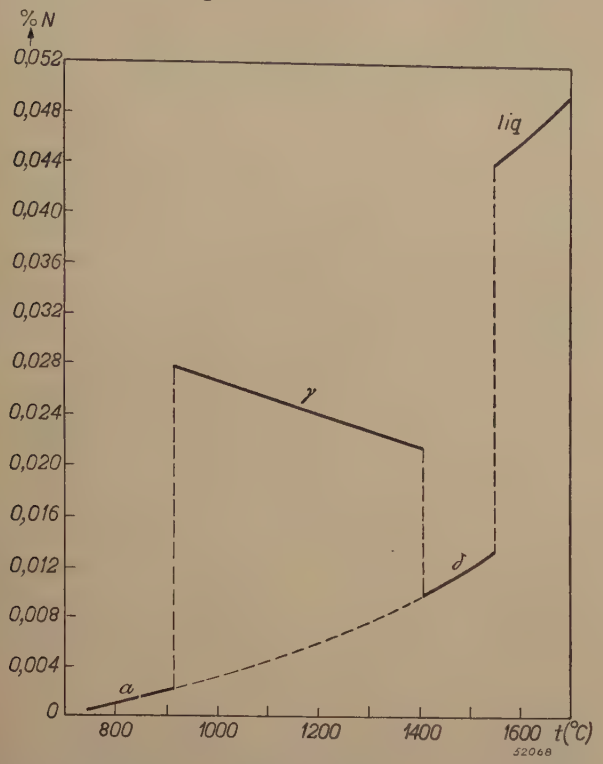



Fig. 4. Solubility of nitrogen in iron at 1 atm nitrogen pressure as function of the temperature. The curves drawn for solid iron correspond to the measurements of Sieverts and his collaborators (see footnote [9])) and for liquid iron to the measurements of Kootz (see note [14])). It is to be noted that for liquid iron various investigators find different values. For instance according to Chipman and Mutphy (Trans. A.I.M.E. **116**, 179-196, 1935) the solubility immediately above the melting point is about 10% lower and according to Sieverts about 30% lower.

can be drawn through the points of alpha and delta iron. This does indeed prove to be the case (see dotted line in fig. 4) and it makes the points for alpha iron more reliable. According to fig. 4 the solubility in the gamma phase, stable between 910 °C and 1400 °C, and the solubility in liquid iron are much greater than that in alpha (delta) iron.

Now in order to be able to say something about the *maximum* solubilities on the grounds of these experimetal data, we shall avail ourselves of the fact that the concentration of a bi-atomic gas in a metal is in general approximately proportional to the square root of the pressure.

From the example of nitrogen this can be realized as follows. The reaction constant $K$ of the dissociation $N_2 \rightleftarrows 2N$ in the gas phase is given by

$$K = p_N^2 / p_{N_2}.$$

The pressure of the atomic nitrogen $p_N$ is thus proportional to the square root of that of the molecular nitrogen $p_{N_2}$. The nitrogen is dissolved in the metal in the atomic state and, since its concentration in iron is small, it will be proportional to $p_N$ and thus also to $\sqrt{p_{N_2}}$. Here, at not very high temperatures $p_{N_2}$ is virtually equal to the total pressure.

Now the temperature coefficient of the solubility in alpha iron is positive, so that the given solubility of 0.0004 % N at 750 °C is certainly an upper limit for the solubility (still at 1 atm) at temperatures lower than 750 °C. For the solubility $C$ (in weight % N) at some other pressure we have

$$C < 0.0004 \sqrt{p_{N_2}} \quad . \quad . \quad . \quad . \quad . \quad (11)$$

An upper limit of maximum solubility is then immediately found by substituting for $p_{N_2}$ the dissociation pressure of $Fe_4N$ as calculated from (10). If it is desired to avoid this pressure, which only arises as a secondary quantity, then one may combine directly the formulae (9) and (11) and write

$$C_{max} < \frac{0.0004\ p_{NH_3}}{\sqrt{K_{NH_3} \cdot p_{H_2}^3}}, \quad . \quad . \quad . \quad . \quad (12)$$

where $p_{NH_3}$ and $p_{H_2}$ are the partial pressures of $NH_3$ and $H_2$ in the gas mixture of 1 atm at which the maximum solubility is reached and the formation of $Fe_4H$ begins.

Thus we find the values given in column 2 of *table III* for the upper limit of maximum solubility.

The values of maximum solubility found experimentally by the various research workers do not by

**Table III.** Values of the upper limit of maximum solubility of nitrogen in alpha iron calculated for various temperatures, and the experimental values of maximum solubility found by various investigations.

| Temp. (in °C) | Maximum solubility (in weight %) | | | |
|---|---|---|---|---|
| | calculated (upper limit) | measured | | |
| 590 | 0.040 | 0.5 [10]) | 0.42 [11]) | 0.13 [12]) |
| 550 | 0.035 | | | 0.05 [12]) |
| 500 | 0.030 | | | |
| 450 | 0.025 | | 0.32 [11]) | |
| 400 | 0.020 | | | 0.02 [13]) |
| 300 | 0.010 | | | 0.01 [13]) |
| 200 | 0.005 | | | 0.005 [13]) |
| 100 | 0.002 | | | 0.001 [13]) |
| 20 | 0.0004 | | | |

[10]) A. Fry, Stahl und Eisen **43**, 1271-1279, 1923.
[11]) O. Eisenhout and E. Kaupp, Z. Elektrochem. **36**, 392-404, 1930.
[12]) D. Séférian, Etude de la formation des nitrures de fer par fusion et du système fer-azote. Paris 1935.
[13]) W. Köster, Arch. Eisenhüttenwes. **3**, 637-658, 1930.

any means agree one with the other, as may be seen from column 3 of this table. Remarkably enough, our calculated values correspond almost completely with Köster's experimental values [13]). Even the lowest values taken from literature are therefore probably to be regarded as an upper limit.

*Solubility of nitrogen in liquid iron and absorption of this element when welding with bare electrodes*

In the foregoing we have considered the solubility of nitrogen only in solid iron. In order to judge the maximum amount of nitrogen that might be absorbed during the process of welding we should have to know the solubility of nitrogen in liquid iron. Considering that welding is usually done in air, at a pressure of 1 atm, the partial pressure of nitrogen then being about 0.8 atm, the solubility at 0.8 atm nitrogen pressure is of particular importance to us.

Various investigators have determined the solubility of nitrogen in liquid iron at 1 atm nitrogen pressure. The highest and possibly the most reliable values were found by Kootz [14]). The solubility $C_{1\text{atm}}$ at 1 atm as a function of the absolute temperature $T$ can be represented, according to his measurements, by the formula

$$\log C_{1\text{atm}} \text{ (in wt. \% N)} = -\frac{1170}{T} - 0.715 \quad . \quad (13)$$

Substituting for $T$ the boiling temperature of iron, about 2750 °K, we find a nitrogen content of 0.072 %. Considering that the nitrogen pressure is only 0.8 atm, this value has to be reduced to $0.072 \sqrt{0.8} = 0.065$ %. For $T < 2750$ °K equation (13) would yield still smaller values of solubility.

Now Losana's measurements (see table I) and also those of other investigators show that when welding with bare electrodes nitrogen may be absorbed in quantities three to four times as much as the maximum quantity (0.065 %) conformable to equation (13).

In our calculations, however, it has been tacitly assumed that the liquid iron is in contact with nitrogen of the same temperature. In point of fact this is certainly not the case, for all available experimental data indicate that temperatures of at least 6000 °K are reached in the welding arc. At this temperature the concentration of atomic nitrogen in the gas phase is already about $10^5$ times as great as that at the boiling point of iron [15]). Taking this into account it is surprising that not a still greater nitrogen content is found when welding with bare electrodes. As we shall see from what follows, in the deposition of the metal more nitrogen is indeed absorbed but this is partly released again in the cooling down and solidifying of the metal.

**Injurious effects of oxygen and nitrogen**

*Porosity*

As already stated, highly porous beads are obtained when welding with bare electrodes. Everywhere in literature we find that this is caused by the formation of CO as a result of the reaction between the oxygen absorbed while welding and the carbon always present in technical iron and steel.

Our foregoing considerations led us to presume that this is only half the truth and that also nitrogen is to be regarded as a cause of porosity. In order to investigate this we prepared an iron entirely free of carbon and made plates of it 10 mm thick and rods 4 mm in diameter. With these carbon-free rods beads were welded on the carbon-free plates and, as was expected, these were found to be almost as porous as the beads obtained when using normal technical wire and plates both containing about 0.1 % C.

Apart from the formation of CO there is obviously another cause of porosity, and this may be assumed to be the release of part of the dissolved nitrogen.

*Mechanical strength*

The mechanical strength of welds is very often judged by measuring the impact value, i.e. the energy required to break in half in one blow a bar of iron of certain dimensions in which a certain notch has been made.

Owing to the great porosity and the high oxygen and nitrogen content (see table I) welds made with bare electrodes have such a low mechanical strength that there is hardly any sense in measuring their impact value. In a subsequent article we shall see that welds made with modern coated electrodes have much lower but still relatively high oxygen and nitrogen contents, the nitrogen content varying from about 0.005 % to about 0.033 % according to the type of electrode and the oxygen content from about 0.03 to about 0.12 %. It appears that a lower nitrogen content is always accompanied by a lower oxygen content and a higher impact value. Since the oxygen and nitrogen contents vary in the same direction it is not possible to determine from tests on welds the effect that each of these elements separately has on the impact value.

We have therefore carried out a number of experiments with iron of the composition of the bare

[14]) T. Kootz, Arch. Eisenhüttenwes. **15**, 77-82, 1941.
[15]) J. D. Fast, Philips Res. Rep. 2, 382-399, 1947.

electrodes, first removing oxygen and nitrogen by repeated high-frequency melting in pure argon and then adding known quantities of one impurity at a time (nitrogen or oxygen) [16]). These experiments have shown that up to a content of 0.033 % nitrogen does not affect the impact value, whereas the addition of oxygen up to the aforementioned content of 0.12 % is accompanied by a gradual lowering of the impact value.

Furthermore, the experiments indicate that the great influence of oxygen on the impact value is due to the fact that this element is present in the form of an oxide which is partly contained along the crystal boundaries. Consequently the metal readily breaks between the crystals.

*Ageing*

Both after rapid cooling and after mechanical working steel is apt to be unstable, gradually increasing in hardness and decreasing in ductilily. One then speaks of "ageing", in the former case "quench ageing" and in the latter case "strain ageing". These processes may be accelerated by heating to 100 or 200 °C.

Ageing after mechanical working is particularly of great technical importance an a great deal has been published on this subject. Many investigators attribute this strain ageing to the presence of oxygen, whilst many others ascribe it to the presence of nitrogen. The cause of the controversy lies in the fact that the experiments have been carried out with technical steels containing both oxygen and nitrogen.

By studying this phenomenon with iron and steel, specially made in this laboratory, not containing both these impurities together, it has been possible to arrive at the definite conclusion that nitrogen causes strain ageing whereas oxygen does not.

Apparently this is related to the fact that nitrogen has a certain solubility in alpha iron (a solubility decreasing with temperature), whereas oxygen is practically insoluble in alpha iron (cf. figs. 1 and 3). Consequently supersaturated solutions of nitrogen in alpha iron may easily be obtained, for instance by rapid cooling from $P$ to $Q$ in fig. 3. If there is a great supersaturation there may be a spontaneous precipitation of finely divided $Fe_4N$ (quench ageing). If there is only slight supersaturation (for instance after comparatively slow cooling) then one can imagine that mechanical working is necessary to start precipitation (strain ageing). It is also possible that in the latter case it is not so much a question of precipitation of a new phase as the occurrence of changes in concentration in the homogeneous solution, which have an adverse effect upon ductility.

As table III shows, the solubility of nitrogen in alpha iron at room temperature is at most a few ten-thousandths per cent and therefore even very small quantities of nitrogen may be expected to cause phenomena of ageing. This is borne out by our experiments.

**Useful effects of oxygen and nitrogen**

*Transmission of heat*

Wyer [17]) has pointed out that heat transfer to the metal is governed to a high degree by dissociation of the gas in the arc, not only when welding in hydrogen but also when welding in air.

Dissociation of the oxygen and nitrogen requires a great deal of energy and consequently the arc voltage is greater than that in a monatomic gas. The atoms recombine for the greater part on the surfaces of the electrodes. thereby releasing again their dissociation energy.

*Transfer of the metal*

Of fundamental importance is the influence that oxygen in combination with carbon exercises upon the transfer of the metal from the electrode to the workpiece.

At first sight it always appears strange that it should be possible to weld "overhead", the droplets being "shot" upwards against the force of gravity. Various investigators have already pointed out that one of the factors playing a part here is the production of gases during the melting of the metal.

To throw more light upon this question we have carried out some experiments with 4 mm wire made of the carbon-free iron mentioned in the section on "Porosity". Overhead welding was found to be impossible with bare electrodes of this kind. Droplets were formed at the end of the electrode but they ran downward along the rod. The experiments were repeated with the same kind of iron but with 0.1 % C added, when it was indeed found possible to obtain an upward transfer of the metal.

We have already seen (table I) that when welding with bare electrodes the C content drops to very low values as a consequence of the reaction with oxygen. The experiments described above give the impression that this reaction leads to small explosions in the melting metal which act as propelling forces upon the drops.

[16]) J. D. Fast, Philips Res. Rep., to appear shortly.

[17]) R. F. Wyer, Gen. Electr. Rev. **42**, 170-172, 1939.

The experiments described seem to show that any other propelling forces which may act besides the development of CO and which have been summed up by Sack [18]) in this journal do not play any important part. These other forces must exercise their influence also when welding with carbon-free rods but they prove to be incapable of overcoming the force of gravity in the transfer of the metal.

Apart from the already mentioned objections attached to welding with bare electrodes, overhead welding with these electrodes, even if they contain carbon, is hardly practicable on account of the fact that owing to the absence of a "cup" much of the metal is thrown off laterally.

If the formation of CO in the metal is indeed the propelling force when welding with bare electrodes, then not only the absence of carbon (even when oxygen is present) but also the absence of oxygen (even when carbon is present) must neutralize the welding action proper. This conclusion is borne out by experiments of Doan and Smith [19]). They welded with bare electrodes in different gas atmospheres. In helium, argon and nitrogen the droplets just fall apparently under the action of gravity. There is no arc blow and no crater is formed (an elliptical depression in the molten or solidified pool underneath the tip of the electrode). Consequently there is little penetration. The addition of only a few per cent of oxygen to the inert gas was sufficient to make the welding normal; the droplets are then driven out with force, a crater is formed and the penetration is normal.

In welding with coated electrodes too the development of gases is fundamental for the transfer of the metal. With the best coated electrodes, however, these gases are not formed by the reaction between oxygen and the carbon in the metal because the coating affords protection against this oxidation. The production of gases then takes place in the coating and, as we shall see in a later article, constitutes one of the most important functions of the coating.

[18]) J. Sack, Philips Techn. Rev. **4**, 9-15, 1939.
[19]) G. E. Doan and M. C. Smith, Welding J. **19**, 110s-116s, 1940.

# ABSTRACTS OF RECENT SCIENTIFIC PUBLICATIONS OF THE N.V. PHILIPS' GLOEILAMPENFABRIEKEN

Reprints of the majority of these papers can be obtained on application to the Administration of the Research Laboratory, Kastanjelaan, Eindhoven, Netherlands. Those papers of which no reprints are available in sufficient number are marked with an asterisk.

**R 42:** W. Elenbaas: influence of cooling conditions on high-pressure discharges (Philips Res. Rep. **2**, 161-170, 1947, No. 3)

The influence of cooling on high-pressure discharges in tubes and in free air is theoretically discussed and for the case of discharges in tubes verified experimentally. The agreement between experiment and theory is satisfactory.

**R 43:** H. C. Hamaker, H. Bruining and A. H. W. Aten Jr.: On the activation of oxide-coated cathodes (Philips Res. Rep. **2**, 171-175, 1947 No. 3)

To obtain a satisfactory emission from an oxide-coated cathode, the degassing of the tube and the various parts inside it must be carried out in a special order and the correct pumping procedure has to be found for each type of tube. One of the factors influencing the cathode-emission has been made the subject of a special investigation which led to the following conclusions: a) glass heated to 400 °C evolves a small amount of hydrochloric acid; b) in a vacuum tube this HCl reacts with the carbonate or the oxide to give $BaCl_2$ or $SrCl_2$; c) when the cathode is subsequently heated these chlorides evaporate and condense on the grid and the anode; d) under electron bombardment the chlorides decompose, thereby producing Cl-atoms or Cl-ions which poison the cathode.

**R 44:** F. A. Kröger: Photoluminescence in the quaternary system $MgWO_4$ - $ZnWO_4$ - $MgMoO_4$ - $ZnMoO_4$ (Philips Res. Rep. **2**, 177-182, 1947, No. 3)

In the quaternary system (Mg, Zn) (W, Mo)$O_4$ four different crystal structures appear. The fuorescence and absorption of products of these structures are studied.

**R 45:** F. A. Kröger: Luminescence of solid solutions of the system $CaMoO_4$ - $PbMoO_4$ and of some other systems. (Philips Res. Rep. **2**, 183-189, 1947, No. 3)

As shown in a previous paper, the fluorescence of solid solutions of (Zn, Mg)$WO_4$ can be interpreted simply as the superposition of the fluorescence of its two components; but in the systems (Ca, Pb) $WO_4$, (Sr, Pb)$WO_4$ and (Ba, Pb)$WO_4$ new emission bands were observed which were attributed to tungstate groups with mixed surroundings of lead and calcium, strontium or barium ions. In this paper the systems (Ca,Sr)$WO_4$, (Ca,Sr)$MoO_4$, and $(Ca,Mg)_3\ WO_6$ are shown to behave as (Zn,Mg)$WO_4$, whereas (Ca,Pb)$MOo_4$ behaves as (Ca,Pb)$WO_4$.

**R 46:** R. Loosjes and H. J. Vink: The i,V-characteristic of the coating of oxide cathodes during short-time thermionic emission. (Philips Res. Rep. **2**, 190-204, 1947, No. 3)

The potential differences existing across an oxide coating during short-time emission (condenser discharge with an "RC-time" of $10^{-4}$ sec) have been measured.

For this purpose a new measuring method was worked out. Using this method it was found that at current densities of about 5-10 A/cm$^2$ remarkable high potential differences exist across the oxide coating (50-200 V) at the normal working temperature (900-1100 °K) at which the experiments were carried out.

**R 47:** J. D. Fast: The reaction between carbon and oxygen in liquid iron. (Philips Res. Rep. **2**, 205-227, 1947, No. 3)

The main subject of this article is the reaction between FeO and C. Assuming a mixture of liquid Fe and liquid FeO to possess a Gibbs entropy of mixing and a van Laar heat of mixing, the activity of FeO in liquid Fe is computed and compared with observed solubilities. Next the activity of C in Fe is derived by comparing the composition of CO-$CO_2$ gas phases in equilibrium with liquid iron and with graphite, using spectrocopic and thermal data. The quotient $(p_{CO} \cdot a_{Fe})/(a_{FeO} \cdot a_C)$ computed from the activities is a constant only if the C-concentration $< 0.1$ percent by weight. This is confirmed by independent arguments.

**R 48:** C. J. Bouwkamp: Calculation of the input impedance of a special antenna. (Philips Res. Rep. **2**, 228-240, 1947, No. 3)

A calculation is given of the input impedance of antenna consisting of a vertical wire fed against a

system of two or four equal horizontal wires. The latter are placed end to end and symmetrically around the base of the antenna proper. The investigation is based on the assumption of sinusoidal current distribution. For a quarter-wave antenna the radiation resistance is found to be approximately 20 ohms, both for two-wire and for four-wire systems.

R 49: F. L. H. M. Stumpers: On a non-linear noise problem (Philips Res. Rep. 2, 241-259, 1947, No. 4)

A rectangular noise spectrum is applied to a valve with a non-linear current-voltage characteristic. The energy frequency spectrum is computed. It is shown that the partial spectra resulting around multiples of the original central frequency have different forms. They are distinguished by their order. If the characteristic is given in the form of a polynomial or of a power series, a formula is obtained fom which all partial spectra can be computed directly. Finally, the presence of one or more carriers is taken into account.

R 50: J. L. Meyering and M. J. Druyvesteyn: Hardening of metals by internal oxidation, II. (Philips Res. Rep. 2, 260-280, 1947, No. 4)

As was described in Part I, certain alloys of Ag, Cu and Ni with a few atomic % of a homogeneously dissolved metal having a suffiicent affinity for oxygen can be dispersion-hardened by diffusing O into them. Two conditions must be satisfied. Firstly the oxide must be formed not as a surface layer but dispersed in the interior of the alloy. In this connection the penetration of the reaction front and the oxide concentration was calculated. Secondly the dispersion must be very fine. The greater the affinity for O of the basic metal, the greater must be the affinity of the solute to produce oxide that conglomerates slowly enough. This was worked out in a tentative thermodynamical scheme.

In Part II diffusion coefficients of O in internally oxidized alloys of Ag and Cu are given. X-ray and and electrical resistivity measurements support the view that the MgO and $Al_2O_3$ particles that harden silver are very small. The mechanical properties are not much affected by long annealings at high temperature. Creep and recrystallization are slowed down considerably. A drawback is the intercrystalline brittleness of these materials, which is less serious when somewhat smaller hardness is aimed at. Single crystals are completely ductile. Some peculiar metallographic effects are explained.

R 51: T. H. Oddie and J. L. Salpeter: Minimum-cost chokes (Philips Res. Rep. 2, 281-312, 1947, No. 4)

A design method is developed for chokes carrying A.C. only, to enable the most economical dimensions to be found for any given electrical characteristics. Formulae and tables are given for rectangular types of chokes with and without limitations on the stacking height. It is also shown briefly how the method may be applied to chokes carrying D.C. with superimposed A.C.

R 52: A. J. Dekker and W. Ch. van Geel: On the amorphous and crystalline oxide layer of aluminium (Philips Res. Rep. 2, 313-319, 1947, No. 4)

Aluminium can be covered electrolytically with a porous layer of aluminium oxide, and oxidation afterwards in boric acid gives rise to the formation of a crystalline layer. The experiments described below show undoubtedly that this layer only fills up the holes of the amorphous $Al_2O_3$. Moreover, there is a correlation between the current density in oxalic acid and the porosity of the amorphous layer thus formed.